NOTHING TO SEE HERE: GENERATIVE AI, NEOLIBERAL CRISIS, AND INTENSIFIED COUNTERINSURGENCY

Grayson James Melville Richards

A dissertation

submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements of the Degree of

Doctor of Philosophy

Graduate Program in Communication and Culture

York University – Toronto, Ontario

November 2025

ABSTRACT

Examining generative artificial intelligence (genAI) as a discursive and technical agent deployed in response to the ongoing crisis of neoliberal legitimacy, this dissertation advances the claim that so-called "safe" models extend the counterinsurgent (COIN) mode of governance, both as an effect of the technology's inherent mode of visualizing "meaning" from unstructured data, and through their operationalization as instruments of epistemic and political management.

Anchored by Mirzoeff's genealogy of Visuality—defined as the 'visualization' of the social in ways that separate subjects and authorize centralized control—the dissertation's narrative is organized into three parts. The first develops a historico-genealogical account of media, visuality, and power, showing how media technologies reflect and intensify particular visualities, culminating in the emergence of algorithms as the media form native to a mode of visualization which 'sees' the social as a heterogeneous field to be stabilized through counterinsurgency. Part two narrates the current intensification as an establishment response to the deepening crisis of neoliberal governmentality and the runaway *aggregation* of insurgency enabled by platform algorithms, securitizing genAI through appeals to epistemic integrity, AI safety, and existential risk in authorization of extrapolitical information controls. Finally, part three traces this intensification of COIN through an analysis of the complementary regulations, technical countermeasures and epistemic interventions constructing the "safe" model as a palliative to the "existential threats" of AI-enhanced disinformation and its corollary: infocalypse.

GenAI emerges at a conjuncture where digital networks have made it increasingly difficult to sustain stability via COIN alone. This dissertation contends that, absent meaningful political economic reformation, authority has instead resorted to progressively heavy-handed interventions in the digital public arena, birthing a complex of protocols which in effect produce "safe" generativity as a means of policing epistemic and political legitimacy through the tactical *disaggregation* of mounting insurgency. Recognizing that, while an extension of COIN visuality, genAI simultaneously reveals this intensification as the product of an ultimately *mutable* system, the dissertation concludes with a speculation on the possibility of appropriating generativity to visualize realities emancipated from the oppressive legacies of Visuality.

Keywords: Artificial intelligence; AI safety; Counterinsurgency; Critical AI; Disinformation; Generative AI; Generativity; New media; Political economy of media; Propaganda; Visuality

DEDICATION

For the future scholars, artists, and activists who,

through brilliance and determination,

will undo the conditions that make this work necessary.

ACKNOWLEDGEMENTS

I am very lucky and honoured to have received the immense support that has allowed me to undertake doctoral work. This research was conducted with funding from numerous institutional and government scholarships, including the York University Graduate Scholarship, Ontario Graduate Scholarship and the Social Sciences and Humanities Research Council Doctoral Fellowship. As a holder of student debt, I want to also extend a special thanks to the public servants, legislators, and activists—past and present—who have fought to make higher education accessible and affordable for working class and resource-poor scholars.

Thank you to the faculty, staff, and communities at York University and Toronto Metropolitan University for their tremendous support and guidance over the years. Thank you to GPAs Trisha Diaram and Sarah Edmonds, and to GPDs Steve Bailey, Jan Hadlaw, and Markus Reisenleitner for their patience and care in helping me navigate the program and the sometimes-labyrinthine universities that house it. Thank you to the colleagues I taught with, especially Cameron Johnston, Paul Baxter, and Myles Wieselman, and thank you to my friend and ComCult alumni Alysse Kushinski, whose early support and handed-down wisdom helped take the edge off an otherwise difficult transition.

I would not have embarked on this journey without the encouragement and inspiration of many past educators and mentors: Brenda Gallander, Arni Haraldsson, Kyla Mallet, Chris Jones, Marian Penner Bancroft, Sandra Semchuk, Vid Ingelevics, Blake Fitzpatrick, Matthew Croombs, Sara Angellucci, Sara Knelman and Ed Slopek, in particular. Thank you also to the artists, curators, writers, and thinkers across the country with whom I've had the pleasure of sharing images and ideas with over the years. I can only hope that my contribution to these communities and relationships moves others the way they've moved me.

I am especially grateful to the friends and family whose love and understanding form the basis of everything I have and will accomplish. Thank you to Nathaniel Brunt, Rory Sharp, Craig Fahner, Kyler Zeleny, Alexa Vachon, Liam Maloney, Mark Dudiak, Jesse Sarkis, Ruth Skinner, Erdem Taşdelen and the others who have listened to many gripes and long, winding, verbal drafts of this dissertation. My parents, Cristine Wierzbicki and Steve Richards, and sister, Caitlin

Richards, for nurturing in me the sense that life is a thing of wonder and contemplation—though not to be taken too seriously—and always letting me find my own way. To Lindsey King, for a love I wouldn't believe was possible if you hadn't shown me, for the spiritual and literal nourishment, and for all the dances.

Finally, I want to express my deepest appreciation for the committee, especially my supervisor Caitlin Fisher, whose dedication and enthusiastic support saw this research through its many twists and turns. Thank you, as well, to David Cecchetto and Marc Couroux, your generosity and brilliance have left a mark on this dissertation, and I aspire to always approach my work with the level of rigor and imagination you have modelled.

TABLE OF CONTENTS

A NOTE ON CITATIONS AND FOOTNOTES

This dissertation follows the Chicago Manual of Style for all citations. Sources are cited using numbered endnotes, which are compiled in the back matter and organized by chapter. This format is intended to improve readability while maintaining easy access to full references in a centralized location. A complete bibliography follows the endnotes.

To preserve the flow of the main text while allowing for occasional elaboration or clarification, footnotes marked with symbols (*, †, ‡, etc.) are used sparingly. These footnotes contain brief discursive comments or supplementary reflections and are distinct from source citations.

A NOTE ON TEMPORALITY AND SCOPE

Given the rapidly evolving nature of the technological and political developments it addresses, this dissertation may reflect different temporal vantage points. While it has largely been written contemporaneously with the objects and events it analyzes, some sections have been revised to account for developments as recently as August 2025. While these targeted revisions reflect an attempt to ensure the accuracy and relevance of key areas, they stop short of retroactively overhauling the entire dissertation to reflect the most current state of affairs. Beyond providing a helpful note for the reader, I feel that an acknowledgment of the dissertation's temporal irregularity similarly illustrates the pace and volatility of the field under study.

LIST OF FIGURES

INTRODUCTION

Prelude

In a funny, roundabout sort of way, when tracing a history of generative AI one might begin with the example of an 18th century machine which rather unambiguously embodies a common metaphor applied to the present state-of-the-art. Atop a wooden base containing the mechanics that allowed it to imitate natural sound and movement, a life-sized duck sat clad in polished copper sheathing, perforated to allow a view of its marquee feature. After eating corn and grain from the hands of visitors, the duck's innards of metal and rubber would come to life, dutifully going to work on the ingested matter for a period of time, after which the gilded duck would shimmy and 'croak' before relieving itself of "an authentic-looking burden."[1] First presented in the hall of a Paris hotel in 1738 by the inventor and artist Jacques Vaucanson, the automata *Canard Digérateur*—or "Digesting Duck"—rests among its contemporaries as an inauguration of the current age's preoccupation with the technical simulation of life and intelligence.

After centuries of experiments which would see this imitation advance from the mechanical programming and performance of 'life' exhibited by Vaucanson's duck, through the randomized simulation of creativity by rudimentary algorithms* and early computational art like

* As in *Musikalisches Würfelspiel*, an 18th century game wherein players "composed" musical arrangements by rolling dice; Gerhard Nierhaus, *Algorithmic Composition: Paradigms of Automated Music Generation* (Springer Science & Business Media, 2009): 36.

that of George Nees and Vera Molnár,* the core processes behind the current state-of-the-art are distinguished by the ability to generate statistically unique data instances without extending or making *specific* reference to any prior pattern. Collectively referred to under the umbrella term 'generative model,' early foundational frameworks like deep belief networks,[2] deep Boltzmann machines,[3] and generative stochastic networks[4] calculated (through varying means) the probability distributions of supplied datasets and 'learned' to create new instances which mimicked the sorts of distributions observed. While the data instances generated by these models were statistically unique, their "learning slow[ed] dramatically after the model ha[d] learned even an approximately correct distribution,"[5] presenting a bottleneck for improvement.

A watershed moment in the history of generative AI came in 2014, when Dr. Ian Goodfellow and his team at the University of Montreal submitted the pre-print article "Generative Adversarial Nets," in which they presented a new framework which enabled generative models to surpass the bottleneck to produce new data of ever-increasing fidelity. Their framework, now familiar to many, trains two models on the same data (a *generative* model producing new patterns intended to resemble the training data, and a *discriminative* model which classifies the generator's outputs as either a 'false' or 'true' instance from the original dataset) and pits them against each other.[6] By incorporating a mechanism for back-propagation, adversarial networks allowed the generator to iteratively adjust its work in response to discriminator feedback until the output is such that the discriminator's classifications are no

* Works like Nees' *Schotter (Gravel)* (1968) and Molnár's *(Dés)Ordres* (1974) are prototypical of early computational art, where artists would provide an initial set a parameters (in these cases, geometric compositions) that simple algorithms would then iterate on, producing novel (and infinite) abstractions or variations on the theme. While generative art of this variety relies on degrees of controlled randomness, and therefore doesn't mark all that significant of a shift from the rolling die of Musikalisches Würfelspiel, their most notable advance is to do with the algorithms' ability to generate new statistical arrangements without requiring that a set of possibilities be predetermined by the programmer.

better than random. Trained with comparatively low resources, the framework created a generative model capable of producing outputs consistently categorized as real (or borderline, at least) by discriminator models and human observers alike. As a groundbreaking innovation, the GAN framework inarguably touched off the last decade of exponential growth and investment in generative AI research. The years following have seen the emergence and rapid improvement of models capable of generating high-fidelity multimodal data, culminating (for now) in the large foundation (or 'general purpose') models like Meta's Llama or OpenAI's GPT-4o, which use a single neural network to interpret and generate "any combination of text, audio, and image."[7]

OpenAI's public release of Dall-E 2 and ChatGPT in 2022 marked another significant turning point in the ongoing history of the technology, democratizing access to the first user-friendly interfaces for producing synthetic media. At the present juncture, despite decades of advancement and the current froth of industry hype, the scatological connotation invoked by the incontinent mechanical duck persists as a popular metaphor for artificial intelligence, where the dictum "garbage in, garbage out" continues to apply as much to the technical workings[8] of a system as it does its sociopolitical effects.[9] As research continues to gather pace and generative models become progressively more capable, accessible, and embedded into the infrastructures and practices of communication and information processing, the technology has become a focal point for a popular imagination increasingly preoccupied with the interrelated threats of epistemic disorder, democratic decline, and existential risk. While the character and usefulness of future developments remain open questions, it is clear that generative AI are a paradigm shifting media, ripe with the double-edged promise of transformation. This dissertation seeks to take up the open question of this promise, presenting a theoretical and genealogical account of generative AI's ongoing development as a political technology, its securitization via the

discourse of disinformation and existential threat, and mapping an emergent complex of regulatory, technical, and epistemic closure fashioning the technology as a means of intensified counterinsurgent control.

Background

Beginning in the later months of 2023 and continuing throughout 2024, much of the political discourse on generative AI was characterized by growing concern about the technology's potential impact on the integrity of democratic elections around the world, particularly in the United States. Dubbed a 'super year' for elections, 72 countries (representing nearly half of the world's population) held elections in 2024.[10] During this time, as many articles,[11] reports,[12] and symposia[13] materialized to announce the imminent threat to democracy, polls similarly reflected growing public anxiety about the deleterious effect uncontrolled use of generative AI might have on already contentious political processes.[14] This threat model (as propagated by many politicians, media, and academics*) regularly cites the low cost and accessibility of leading generative models as a major contributing factor to increased risks of electoral disruption, naming the generation of high-volume, micro-targeted, and strategically-timed disinformation as a chief concern.[15] Echoing identical narratives[16] about these effects disrupting the 2020 presidential election,[17] the ongoing panic over AI-enhanced disinformation

* Hany Farid, Professor at UC Berkeley's School of Information and regular commentator on the sociopolitical threat of generative AI, maintained a website dedicated to the tracking and analysis of deepfakes related to the 2024 US Presidential election; Hany Farid, "Deepfakes in the 2024 US Presidential Election," accessed November 26, 2024, https://farid.berkeley.edu/deepfakes2024election/.

has led to a slew of state-level laws, national and regional regulatory frameworks, and a stream of government appeals requesting that leading AI companies develop guardrails preventing their products from being used to campaign, influence public opinion, or otherwise interfere with political processes.[18] While the anticipated disruption to 2024 elections never fully materialized,[19] the situation is exemplary of a perpetually reinvigorated discourse on generative AI's cataclysmic repercussions, where repeated lack of evidence or result simply defers the threat onto futures where the technology is more advanced.*

From the time they first entered the public consciousness, generative AI and the synthetic media they create have been consistently discussed as representing an existential threat to 'objective reality' and the liberal democratic order alike; a claim which rests on a narrative foundation which implicitly (and often explicitly[20]) takes the two as synonymous. As such, much of the mainstream political discourse around generative AI has produced a threat model wherein the technology becomes inseparably intertwined with the wider phenomena of online disinformation and the erosion of trust in establishment† institutions being observed across the liberal democratic world, particularly in the United States, United Kingdom, and Canada. Under this model, the existence of and public access to high-performing generative models commonly stands in as a cause and accelerant of disinformation and worsening epistemic disorder, leading to increased political polarization, rising populism, and the degradation of the 'information environment' as a vital pillar of stability and resilience in open societies.[21] Faced with this

* Speaking to Time Magazine about the technology's lack of effect on 2024 elections, vice president of political affairs at Encode Justice, Sunny Gandhi, reasoned that current models simply weren't capable enough to materialize the threat, saying "I'm pretty worried about what it will look like in 2026 and definitely 2028.": Ibid.

† In the context of this dissertation, the term "establishment" is invoked in reference to the global complex of—predominantly centrist— institutions (political, economic, media, etc.) with a strong stake in the existing social and political order.

multivalent threat, institutions of global neoliberal governance (i.e. those invested in what international relations refer to as the 'liberal international order') have developed a general understanding of the current moment as presenting a narrow window of opportunity for redress, resulting in a rush to "implement robust alternative countermeasures to address this growing danger"[22] and establish the codes and controls of the future "generative world order."[23] As it develops, this project has adopted the concept of 'safe and responsible' AI as the organizing principle central to an emergent complex of regulatory, technical, and epistemic closure directed at curtailing the technology's destabilizing effects.

Simultaneously a discourse and a field of practical research, 'AI safety' refers to both the social and philosophical deliberation of AI's potential risks and the development of technical and regulatory mechanisms for their mitigation, with a focus on ensuring that present and future artificial intelligences 'align' with human values and ethical standards.[24] Within the discipline, the goal of alignment is typically understood to be the avoidance of catastrophic *existential* risk and the lesser, but still significant, *systemic* risks resulting from the deployment of misaligned or "unsafe" systems. In this context, the term 'existential risk' principally references the potential of advanced AI—specifically artificial general intelligence (AGI), or *superintelligence*[*]—to bring about the annihilation or irreversible collapse of humanity through *mis*alignment, whereas 'systemic risk' refers to their potential to adversely effect core societal infrastructures and outcomes.[25] On the whole, the field strives to develop technical methods, standards of practice,

[*] Setting it apart from AGI (as human-like intelligence) Nick Bostrom defines superintelligence in his eponymous 2014 book as the emergence of "machine brains that surpass human brains in general intelligence," and therefore especially dangerous.; Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, First edition (Oxford: Oxford University Press, 2014): vii.

and other safeguards to ensure that present and future AI systems are controllable, aligned, or otherwise 'safe' for real-world application.

Despite their regular citation as an intermediary step to AGI,[26] the specific risks posed by current generative models largely remain confined to the realm of systemic harm.* While this capacity for harm is perhaps most often reflected in the exacerbation of structural inequalities resulting from the tendency of models to reproduce bias and discrimination when deployed in public-facing contexts,[27] generative AI also have the potential to displace human labour through the automation of cognitive and creative tasks, expose private information collected in the course of their training or operation, and/or facilitate cyberattacks or other interference with critical infrastructure. In each of these instances, potential for systemic risk develops out of a quantifiable technical limitation inherent to the AI system, and/or emerges as an effect of its negligent or malicious deployment. Notably, while discussion of disinformation as an *existential* risk exacerbated by 'unsafe' generative AI remains largely absent from primary AI safety literature, it has emerged as a point of focus in secondary institutional and political discourses, which often invoke the wider discipline of AI safety in order to support or justify claims about the need to secure the digital public sphere against the threat of AI.

While political assertions regarding the disruptive systemic effects of disinformation† are relatively uncontroversial, I contend that discourses framing it as *existential* risk necessarily hinge on a problematic conflation of 'safety' (and 'existence,' for that matter) with the ongoing stability and persistence of incumbent regimes of discursive and political economic power.

---

* Barring their contribution to climate change as a result of increasing resource-consumption.

† The difficulty of defining disinformation, and the resulting political and methodological implications, is more fully addressed in the literature review.

Applied to generative AI, such existential rhetoric is readily observable in reports like "The Existential Threat of AI-Enhanced Disinformation Operations" from the Center for Security and Emerging Technology,[28] a think tank based at Georgetown University, and the early popularization of the term "infocalypse" (a portmanteau of 'information' and 'apocalypse' that does anything but understate its stakes) to describe the "fucked up dystopia" of a future information ecosystem upended by generative AI and other decentralizing media technologies.[29] Central to this framing is the presentation of generative AI—particularly in their application as communication technologies—as an unprecedented threat to the "epistemic security" of liberal democracies, defined in terms of the information infrastructure's resilience to "adverse influence"—a category that, notably, extends beyond foreign interference to include domestic dissent and the use of new communications technologies by anti-establishment or otherwise insurgent political movements.[30] Following this, I contend that the linkage of generative AI to disinformation and the erosion of the liberal democratic order is effectively the discursive product of AI safety's *politicization*, a result of establishment institutions and allied political actors broadening the definition of existential risk such that "safe" AI development has come to also entail the preclusion of potential social and political transformation resulting from the technology's "misuse."

Insofar as they are a significant factor in the developing political economy of synthetic media, I believe we must consider the historico-political context into which discourses espousing

the risks of generative AI are emerging—namely, the crisis of legitimacy currently affecting the norms and institutions of neoliberal governmentality. While accounts vary as to the origin of the present conjuncture, the sub-prime loan crash of 2008 and the following 'great recession' are now generally understood as marking the moment when the already-tenuous legitimacy of neoliberalism, which had for years "provided ideological cover"[31] for rampant privatization and the unfettered accumulation of wealth, effectively collapsed.[32] Despite being the direct result of unchecked deregulation, policy responses* to the financial crisis largely functioned to 'prop up' a global economic system still organized around core tenets of the neoliberal ideology so recently and fantastically undressed.† While the bailouts and temporary stimulus of this 'administered neoliberalism' successfully avoided Great Depression-level devastation, the years following the crash have largely been characterized by austerity, worsening inequality, and ballooning government deficits despite continuation of deregulatory policy and the privatization of public services. In the early 2010's, absent meaningful reform, neoliberal legitimacy continued to falter as media conditions shifted and the disparity between rhetorics of 'recovery' and actual policy outcomes became increasingly difficult to reconcile. As a result, a number of anti-establishment populist movements began to take shape around the shared discontents of citizens from around the world and across the political spectrum.

In September 2011, inspired by the Tunisian protesters who took to the streets to protest the corruption and economic stagnation experienced under the 24-year rule of Zine El Abidine

---

* I specify policy responses because, at the time, there was a brief period of introspection, as some Western leaders publicly questioned the neoliberal orthodoxy and its role in the crisis.

† While there was some variance in remedy on an international level, the Emergency Economic Stabilization Act of 2008 enacted by the US Congress is emblematic of the establishment response, earmarking $700 billion of taxpayer money to purchase "troubled assets" from a handful of financial institutions deemed "too big to fail." Congress.gov. "Text - H.R.1424 - 110th Congress (2007-2008)[…]." October 3, 2008. https://www.congress.gov/bill/110th-congress/house-bill/1424/text.

Ben Ali, a group of activists in the United States—protesting growing inequality and the perceived capture of the democratic process by corporate interests—set up an encampment in New York City's Zuccotti Park. By the end of that month, similar protest movements appeared in over 900 cities around the world.[33] In both the Arab Spring and the worldwide Occupy movement that it inspired, the combination of widely-accessible internet-connected mobile devices and decentralized social platforms like Twitter and Facebook allowed activists to organize and share information at unprecedented speed and scale.[34] Through the appropriation of new communication technologies, previously atomized citizens were able to quickly and effectively "constitute themselves as a collective actor," building solidarity and mobilizing a non-hierarchical political movement around shared disaffections with a neoliberal status quo characterized by deregulation, privatization, and growing income inequality.[35] Interestingly, while populist uprisings in North Africa and the Middle East were largely welcomed by the western political establishment—which hoped revolution would bring democracy and greater global economic integration to the region—the vast majority of Occupy encampments in western centers were forcibly removed within 3 months of the movement beginning.[36] As has since become apparent, the swift repression of the popular political sentiments represented in the left-leaning Occupy Movement would in turn create the conditions for the festering and later reemergence of new populist challenges to the global neoliberal establishment.

In the years following Occupy, the (still) mostly unregulated social media platforms continued to decentralize narrative/communicative power, circumventing traditional information controls[37] and effecting the algorithmic clustering and connection of like-minded users.[38] Widely understood as populist repudiations of the political economic status quo, the electoral upsets of Brexit and the 2016 US presidential election again demonstrated social media's ability

to foster the *aggregation* of previously fragmented publics into collectives of political consequence, with social media playing a central role in the strategies of the anti-establishment #Leave and #MAGA campaigns. Correlating trends toward decentralization in the public sphere with the realization of undesirable electoral outcomes, establishment narratives have since explained this drastic shift in public sentiment by mostly eschewing the destabilizing effects of rampant political and economic marginalization, instead attributing the legitimacy crisis to changing media conditions and a collective lapse of judgment on part of easily-manipulated "resource poor" citizens.[39] The core assumptions of this narrative would later become formalized in the post-2016 ascendance of 'disinformation' as a subject of urgent political and academic concern.

In a blog post designating *post-truth* as its 2016 'Word of the Year,' Oxford Dictionaries explicitly referenced both Brexit and Donald Trump's election as exemplary of the growing influence of "emotion and personal belief" over "objective facts" in shaping public opinion.[40] Around the same time, well-funded researchers in the newly-formed discipline of 'disinformation studies' took up the task of explaining the dynamics behind the rise of 'post-truth politics.' By and large, this work functioned to quantify the increase in anti-establishment sentiment as the result of a flawed information ecosystem, producing research which in turn gave credence to the de-politicized view of widespread public disaffection "as an issue of information quality and not an expression of structural tensions and transformations."[41] Translated to public discourse, these narratives also increasingly employed the language of war as a primary frame for representing online information dynamics, conflating the circulation of dissenting and alternative narratives more broadly with the 'weaponization' of social media; an existential threat to be 'defeated' under the banner of the 'war on disinformation.' As generative models have

become more available and concerns about the political effects of AI-enhanced disinformation increase, martial rhetoric on the 'fight' against disinformation has expanded to the new technology, with tech companies, policymakers, and media pundits alike routinely expressing the need to "battle,"[42] "combat,"[43] or "counter"[44] the rising threat. Mistaking the symptom for the cause, establishment disinformation narratives transform sociopolitical dissension into the effect of a misaligned technical system, creating a tendency among government and industry policymakers to privilege "narrowly technical and technocratic solutions"[45] when addressing the crisis.

According to some critical scholars, the disinformation narrative's adoption of martial language and a demonstrated[46] preference for technical resolutions to the crisis of legitimacy—a decidedly social and political phenomena—mirrors the strategic misadventures of earlier establishment attempts at social and political relegitimation: "We are repeating the mistakes of the war on terror, prioritizing repressive, technologically deterministic solutions while failing to redress the root sociopolitical grievances that cultivate our receptivity to [dis]information in the first place."[47] In the early years of the global "war on terror" the United States and its 'coalition of the willing' enacted a particularly repressive strategy of counterinsurgency (COIN), a military doctrine which seeks to re-legitimate a beleaguered establishment through the forceful elimination or separation of the intransigent 'insurgent' from a 'host' population otherwise amenable to its authority.* Critical perspectives on disinformation suggest that the theorization

* COIN theorist David Kilcullen broadly defines *counterinsurgency* as "the complete range of measures that governments take to defeat insurgencies." This follows from the US military field manual's definition of *insurgency* as "an organized, protracted politico-military struggle designed to weaken the control and legitimacy of an established government, occupying power, or other political authority while increasing insurgent control." Insurgencies which fit this definition are typically made up of 'irregular forces,' that is, they operate outside of inter-state military conventions; David Kilcullen, *Counterinsurgency* (Oxford University Press, USA, 2010): 154.

and prosecution of a war on insurgent information reveals a similar tendency within the establishment to see its declining legitimacy as the effect of a technical and political irrationality that must be *suppressed*, rather than the symptom of deeply-rooted grievances that are at once rational and in need of redress.[48]

A Note on Counterinsurgency

In the context of this dissertation, rather than marking general opposition or serving as a catch-all for establishment or otherwise conservative politics, 'counterinsurgency' names a specific *mode of governance* tailored to conditions of contested legitimacy. In contrast to other modes, counterinsurgent governance sustains authority and seeks the consent of the governed not through political/ideological responsiveness and renewal, but through techniques of *stabilization*—identifying, separating, and (ideally) neutralizing real and perceived threats to the regime of legitimate authority.[49] Though most clearly legible in military and foreign policy, counterinsurgency increasingly permeates the domestic spaces of neoliberal governmentality—a trend intensified by recent spikes in populist unrest and the rise of algorithmic surveillance and control. Distinguished from establishment politics in a general sense, counterinsurgency names that rationality which is at play when authority, unable to maintain legitimacy through normative political means, turns to stabilization as the means of (and justification for) its mobilization of political, epistemic, and physical force. As such, it marks the mode of governance whereby a self-legitimising authority achieves and maintains stability not by addressing the political roots

of insurgency, but by rendering it invisible, disaggregated, or otherwise inoperable through tactical means.

As I argue in the literature review—and later illustrate in the dissertation's coda—counterinsurgency is best understood as an operational logic (inherent to machine learning and neoliberal governance insofar as they are both cybernetic systems) which is *functionally independent* of its various political, epistemic, and strategic deployments. This point is particularly salient with respect to our thinking about generative AI, where even well-meaning efforts to de-bias or otherwise 'decolonize' generative models can only hope to redirect and reproduce counterinsurgency so long as they preserve the underlying operational principles of AI knowledge production and generativity. Counterinsurgency, then, should be thought of as *apolitical* insofar as it substitutes the technical for the political in mounting a response to the problem of insurgency, however defined.[*]

Counterinsurgency is further illuminated by clarifying another concept deployed throughout this dissertation: homeostasis. Originating in physiology, homeostasis describes the tendency of self-regulating biological systems to maintain optimal internal conditions in spite of external variables. This state of internal equilibrium is created and stabilized through various mechanisms of negative feedback—that is, the detection and counteraction of perturbances through compensatory actions. A simple biological example of such feedback is the activation of sweat glands in response to rising temperatures, an external deviation which the body autonomously counteracts through the application of negative feedback in the form of

[*] As with counterinsurgency, this dissertation adopts a definition of the 'insurgent' that, beyond its commonplace military association, is more broadly inclusive of the various aggregates of discontent, political force, and epistemic excess that disrupt order and elude containment.

evaporative cooling. Being derived from the Greek *homoios* (similar) and *stasis* (standing still), the etymology of the concept further underscores its character as the chief regulatory principle for systems with the goal of 'staying the same'.

Homeostasis is also central in cybernetics, which applies the concept of the feedback loop to understand the regulation of more complex systems made up of biological, technical, and/or social elements. In the context of governance, as it's applied here, it names the operational goal of fundamentally *reactive* political logics which seek to stabilize an incumbent order through the monitoring and neutralization of deviant (or insurgent) subjects and behaviours. Homeostasis thus not only describes an operational aim of counterinsurgency (which imagines stability as the means and ends of its legitimacy—itself a circular logic) but also neoliberal governance more broadly, as political authority and technical systems come to reflect a shared preoccupation with stabilization as the operational antecedent to continued market expansion and administrative control.

The fundamental reactivity of homeostatic logic is distinct from other modes of cybernetic stasis that, while also concerned with regulation, differ in their framing of the relationship between deviance and stability. For example, the concept of *allostasis* describes the maintenance of stability through the pre-emptive counteraction of anticipated stressors. Whereas homeostats are relatively simple mechanisms, allostatic regulation requires the tracking and comparison of multiple variables over time in order to predict needs and modulate the internal conditions of a system in advance of its disturbance.[50] While it is often understood as more adaptive, in the context of governance, allostasis still describes a set of regulatory mechanisms committed to the stabilization of an incumbent state through the management of externalities, and as such reflects an idealized horizon of counterinsurgent control. As the concept

operationalized by anticipatory applications of AI like predictive policing, counterterrorism, behavioural marketing and trend forecasting, allostasis underpins the neoliberal dream of a perfectly administered future where technical and political systems are capable of altogether pre-empting disruption by appropriating the 'noise' of insurgency "as a risky and unforeseen opportunity to perfect the system's survival."[51]

In contrast to both homeo- and allo-stasis, the concept of *metastability* offers an approach for treating instability as a condition of possibility rather than risk. According to the framework offered by Gilbert Simondon, metastability maintains a state of internal disequilibrium from which new forms and structures emerge, theorizing stability as the grounds for securing a system's own "coming-into-being, rather than remaining perpetually in the same state."[52] Whereas the homeo/allostatic logics of counterinsurgency assume the "primacy of being over becoming,"[53] Simondon's metastability opens onto a politics of progress where, rather than simply being preserved, stability sets the stage for the continuous renegotiation and reformation of legitimacy. With respect to the argument put forward in this dissertation, the political stakes of specific regimes of AI governance are thus found through an interrogation of the regulatory logics they encode, constructing the technology as mechanisms of homeo/allostatic stabilization, or fostering them as powerful sites of political and epistemic reimagination.

In the context of the present crisis, then, it comes as no surprise that counterinsurgency should be so deeply embedded in the discursive logic of an ideology which, despite its repeated failures and weakening electoral support, continues to understand itself as the end of history. As changes in media conditions precipitate the unprecedented aggregation of insurgency—first as Occupy, later in its resurgence as a caustic right-wing populism—it becomes increasingly

apparent that the necessary redress might entail nothing short of the order's undoing. Having passed a "tipping point of irrecuperability,"[54] the establishment thus seeks legitimacy through more repressive efforts, doubling-down on the post-political administration of political and economic affairs while intensifying counterinsurgency as the technocratic management style best suited to exercising authority within an increasingly anarchic political terrain.

Perhaps the most salient product of the disinformation narrative has been the buildup of a new counterinsurgent apparatus specifically tailored to the decentralized domain of online information circulation. The emergence of this apparatus—reflecting the technological determinism of a dispersed 'coalition of the willing' from across government, civil society, and media which both authorizes it and puts it to work—portends nothing short of a mobilization of the platform internet *against* insurgency, making the exercise of counterinsurgency coextensive with its technical operation. Having done an about-face[55] after originally resisting[56] their characterization as a leading contributor to the crisis, social media companies have since become a leading edge in the fight against disinformation and insurgent aggregation on their platforms, installing measures to remove content and accounts which violate community standards,[57] algorithmically reduce the visibility and spread of 'false' or 'inauthentic' information under the mantra "freedom of speech, not freedom of reach,"[58] and labeling certain content or accounts as politically or epistemically suspect.[59] In this way, platforms have emerged as central actors in the increasingly martialized space of digital communication and connection, engaged in politically-inflected forms of deletion, demotion, and deterrence. As media scholar Jack Bratich has argued, such measures amount to "acts of war on the digital battlefield," and thus signal the platform's enrollment as de facto *dissent managers* within the broadening counterinsurgent alliance.[60]

Today, leading AI companies—under increasing pressure from a 'coalition' to which many of them already belong*—have also become allies in the fight against disinformation, enacting usage policies and technical mitigations intended to discourage or prevent 'misuse' of their products. Positioned, as they are,† to become ubiquitous mediators of digital communication, generative models have the potential to act as particularly potent vectors of counterinsurgent control, capable of restricting and/or modulating the content and flow of information across a growing range of platforms and devices. Similar to traditional platform services, AI companies employ content/usage moderation (often using secondary classification models) to monitor and block specific user requests or model outputs deemed to be "potentially harmful."[61] Companies also erect guardrails tailored to specific social or political contexts, as in the run up to the 2024 US presidential election when OpenAI,[62] Anthropic,[63] Google,[64] and Meta[65] each modified their systems to elevate "authoritative sources of information," block political content, and disrupt the ability of ill-defined "threat actors" to "build sustained audiences." In addition to the already well-established practice of content moderation, generative models—through complementary processes of training and alignment—further extend counterinsurgency as both an inherent and intentional political/epistemic operation, reflected not only in the statistical operations that enable 'generativity' in the first place, but in the later development of methods for filtering training data and/or adjusting the structure of models' semantic space to suppress or altogether *erase* "undesirable concepts" from their knowledge

* By virtue of owning popular platforms which predate their foray into generative models, companies like Meta and Google have already "signed on" as allies in the fight against online disinformation. These pre-existing commitments generally extend to the development and integration of generative AI into their existing services.

† In a 2023 keynote, NVIDIA CEO Jensen Huang presented the vision of a future where generative models are "on the front end of just about everything" we do with computers. NVIDIA. "NVIDIA Keynote at SIGGRAPH 2023," Youtube. Aug. 8 2023. 20:00-20:42. https://www.youtube.com/watc,h?v=Z2VBKerS63A.

base.[66] Given the prominence of discourses linking generative AI to the existential threat of insurgent communications, these and other measures for ensuring their 'safe and responsible' development as a "force for good"[67] risk capturing and further constructing the model as a powerful means of epistemic and political regulation, suppressing dissent and policing the borders of legitimacy against threat. As will be discussed at a number of points in the dissertation, this trend not only survives, but is exacerbated by the recent political shifts taking place in liberal democratic countries around the world, the United States in particular.

To function reliably, generative AI *must* circumscribe and enforce a regime of legitimacy. Essentially, generative models produce new media by 'learning' and replicating the probabilistic distributions of visual or semantic relationships observed in their training data. As the initial training process is often 'unsupervised' (meaning the model 'learns' associations without human guidance or instruction) the behaviour of so-called 'pre-trained' models tends to be overly general and prone to the reproduction of errant patterns. In order to further 'steer' this behaviour, secondary processes have been developed that direct pre-trained models to categorize patterns as either desirable (to be replicated) or undesirable (to be suppressed), thus internalizing a set of rules or conventions for functioning reliably within a given context. As what's produced through the process of training and alignment, this 'regime' of legitimate patterns actively shapes model behaviours to be both internally *consistent* and externally *acceptable* in terms of adherence to predefined norms. Processes of 'fine-tuning' and 'alignment' are mostly uncontroversial with regards to generative AI's application within fields such as healthcare, systems management, or customer service, where intentional misuse, hallucination, and other 'undesirable' behaviours can affect outcomes in ways that are both unambiguous and immediately consequential.[68] The practice becomes fraught, however, in the context of generative models' deployment as

communication technologies, where the definition and technical enforcement of ‘acceptability’ as it relates to communication and discourse inevitably undermines core democratic values of free speech and unfettered access to information.

Following this, I suggest that principles of AI safety are *misapplied* to generative models in their use as communication technologies,* where distinctions between ‘safe’ and ‘unsafe’ hinge on secondary determinations of what constitutes ‘good’ and ‘bad’ information or patterns of interaction between users, often with respect to subjective matters of culture or politics. Read in the context of neoliberal crisis and the ‘war’ against disinformation, safety measures targeting bias, discrimination and other easily defensible content restrictions can then function as a fig leaf for intensifications of counterinsurgent control. While this is not to suggest that hateful, discriminatory, and deceptive communications or generations are not antisocial or undesirable, efforts to assuage social and political ills by making their expression a *technical impossibility* simply extend the suppressive logic of counterinsurgency, authorize broad information controls, and further deny the right to look in the name of ‘safe and responsible’ development.

Expanding the criteria for its determination, political invocations of ‘existential risk’ similarly authorize the creation of new controls specific to the conjoined threats of disinformation and insurgent aggregation. The mainstreaming of discourses linking disinformation, generative AI, and existential risk bears all the hallmarks of the *securitizing move*; a concept from international relations referring to an issue’s rhetorical presentation as being existential in an attempt to create the sense that, if not dealt with urgently, it will weaken the host society such that future remedy would be either ineffective or impossible, and thereby

* Specifically referring to generative models mediating user interactions with information and/or other human users.

requires emergency action falling outside the normal bounds of political procedure.[69] Conflating 'society' with the establishment, the existential rhetoric around generative AI effectively *securitizes* the new technology, leading to countermeasures to install 'stability' as a goal of the aligned technical system. So, it follows that, under conditions of insurgency, the development of 'safe' AI entails the suppression of destabilizing patterns, a counterinsurgent operation executed by models fine-tuned to inhibit the reproduction of potentially transformative political or ideological formations—extra-political control in breach of otherwise inviolable rights.

The intensification of COIN is further enabled by the growing quasi-governmental power of tech companies, which administer an increasing array of core social services like transportation, healthcare, and communications. While most liberal democracies have provisions protecting freedom of expression, these rules generally only apply to the direct actions of government. Despite this, growing regulatory pressures (resulting from the coalition's amplification of disinformation narratives) to practice 'safe and responsible' AI development effectively transfer the task of conducting COIN onto private companies free to moderate how and what users communicate using their platforms, a process of moral suasion legal experts call "jawboning."[70] Leaving aside the argument that allegations of jawboning are complicated by the parties' shared interests, as more of public discourse takes place on platforms and through other privately-owned channels, the unbridled intensification of COIN is able to proceed under cover of corporate responsibility and the right to enforce terms of use.

In the months following the initial drafting of this dissertation, significant electoral developments in the United States have meaningfully altered the global political landscape with respect to concerns of disinformation and the governance of artificial intelligence. At the time of this revision (August 2025), a growing schism has developed between the discourses on AI

governance emanating from the United States under President Donald Trump and the safety-oriented frameworks still dominant in the European Union and other liberal states. The dissertation has been revised at key points to account for these changes, in particular the latter's shift from 'safety' and regulation to 'security' and *deregulation* as the frames best suited to authorizing its intensification of control, though now in service of enforcing and policing a new—illiberal—regime of legitimacy.[*] The resulting situation is perhaps most clearly articulated in the dissertation's coda, where the landscape of global AI governance is read as a geopolitical contest between counterinsurgent visualities, with generativity deployed as a means of exporting mutually-exclusive regimes of political and epistemological legitimacy. Ultimately, while these changes reflect a (temporary) reorientation away from strategies of incumbent stabilization toward more offensive operations, I argue that they all but underscore the continued intensification of counterinsurgency as the primary discursive frame of visualization and governance.

The neoliberal 'order of things' is in crisis. Its capacity to discursively legitimize its authority is in rapid decline. So impoverished of vision that its narratives of change have become constrained to the diversification of power around an otherwise unperturbed centre, its only recourse to right-wing populism and the increasing aggregation of anti-establishment insurgency is a doubling down of control. The political and economic disillusion of its publics—a call for reimagination and redress—is being met with the further intensification of a counterinsurgent apparatus central to its authority since inception.[71] The securitization of the digital

[*] See "Regulation and AI Safety/Security Discourse" under sub-heading *Regulatory Closure*; *Generative AI, Counterinsurgency, and Intensification*; and the coda, *Hegemony and Threat: US/China, Generative AI and the Contest of COIN Visualities*.

communications infrastructure, scapegoated by narratives holding it liable for increased distrust and the breakdown of discursive hegemony, authorizes new complexes of closure making counterinsurgency coextensive with communication in the extended political space of neoliberal governmentality. In light of how this background comes to bear on the ongoing development and deployment of generative AI as a mediating interface for digital communication, this dissertation seeks to present and support the speculative claim that 'safe' generative models are positioned to become central to an intensified apparatus of counterinsurgent control—presaging a new front in the establishment's wider "war of restoration."*

Dissertation Summary

Having presented the claim that, as it is currently being developed, generative AI represents an intensification of counterinsurgent control—an effect of its securitization via the politicized discourse of disinformation, safety, and existential risk—this dissertation seeks to trace a narrative of this development across historical, political, and technical registers. Deploying a framework anchored by Nicholas Mirzoeff's genealogy of visuality[72]—which I apply to an analysis of media technologies vis-a-vis the non-representational practice of *visualizing* the social in ways that legitimize centralized authority—the dissertation locates generative AI at a point where media decentralization has made the authority of counterinsurgent

* Linking the current conjuncture to the history of armed conflicts aimed at re-establishing the authority of a previous political, social, or religious order, Jack Bratich characterizes the neoliberal response to crisis as "a war of restoration, whose goal is a unification of […] society around a political spectrum dominated by a center."; Bratich, "Civil Society […],": 312.

visuality (as the presently dominant mode of visualizing and authorizing establishment power, whereby authority 'sees' the social as a heterogeneous field to be managed) increasingly difficult to sustain, leading to its intensification via discursive and technical interventions in the digital public arena. The resulting regulatory, technical, and epistemic closures currently forming around generative AI are thus presented as an effect of this intensification, forming a complex of protocols which produce the 'safe' generative model as a means of both tactically *disaggregating* insurgency and discursively shoring up incumbent regimes of legitimacy. Finally, the conclusion summarizes the dissertation's argument and emphasizes the stakes of generative AI's capture as COIN intensification, before speculating on the possibility of appropriating generative AI to visualize social and political realities emancipated from the discursive legacies of visuality.

To construct this narrative, the dissertation is organized into three parts: a genealogical account of media, visuality, and power (Visuality); an analysis of the multivalent crisis as point of intensification (Crisis); and an account of the "safe" generative model as the both the product and agent of an intensified counterinsurgent visuality (Intensification). In part one, Nicholas Mirzoeff's concept and genealogy of visuality is introduced as the core theoretical framework for analyzing how specific media technologies function to authorize, challenge, and/or intensify particular regimes of visualization and power. In Mirzoeff's formulation, visuality* refers to the practice of—and an exclusive claim to—visualizing history and the social in ways which supplement and naturalize the violent separations required of centralized authority.[73] Described as a "discursive practice that has material effects," visualization is comprised of three steps:

* A clarifying note: this dissertation does not capitalize *visuality* when referring to the concept in general terms, or if it is accompanied by a qualifier which otherwise specifies its object (i.e. imperial, counterinsurgent, etc.). Alternatively, it is capitalized as *Visuality* when used as a proper noun, referring to the specific complex of visualization that has historically worked to authorize centralized and/or autocratic leadership.

classification, separation, and aestheticization. Taken together, these form what Mirzoeff calls a *complex of visuality*, a historically specific set of complementary discursive operations which produce the "visualized deployment of bodies and a training of minds, organized so as to sustain both physical segregation between rulers and ruled, and mental compliance with those arrangements."[74]

As the primary goal of visuality is to naturalize separation and prevent those visualized from "cohering as political subjects,"* Mirzoeff constructs a genealogy indexing the progressive adaptation, or *intensification*,† of visuality's discursive operations in response to the constant evolutionary pressure of the subaltern's *counter*visualization as a collective, expressed through dissensus with the "existing realities" visualized by authority and the performative claim to autonomous self-visualization, what he calls *the right to look*.[75] Using this genealogy as an anchor—which Mirzoeff breaks down into three successive 'complexes' of visuality—part one is structured around a periodization correlating the complexes of *plantation*, *imperial*, and *counterinsurgent* visuality to parallel genealogies of power (i.e. *sovereign*, *discipline*, and *control*) and media (*inscription*, *index*, and *data*) to consider how technologies of representation both compliment and intensify the exercise of discursive and material power within each period by enabling new means of visualizing and ordering the social. Following a brief explication of inscription (which naturalized the sovereign authority of plantation visuality through the aestheticizations of natural history, force of law, and the colonial map) and the index (where the 'objective' technology of photography played a central role in naturalizing the authority of

* In this sense, all visuality is fundamentally *disaggregative*, if not always counterinsurgent.

† In the Foucauldian sense, where *intensité* describes a tendency for power to become lighter and more economical in its deployment of force.

imperial discipline and the supposedly 'scientific' classification of 'primitive' or 'deviant' subjects and behaviours), data is then presented as the medium native to the discursive and material operations of counterinsurgent visuality.

An intensification of visuality through a confluence with mechanisms of Control, counterinsurgency diverges from the imperial visualization and reform of non-conforming subjects to re-imagine the social as a heterogeneous field to be *managed* through techniques of homeostatic modulation, regulating the terrains of culture and information so as to produce legitimacy* as an effect of its control. Following this, the section presents a reading of the internet and platform algorithms as political technologies which perform counterinsurgency through the technical protocols governing their operation, circumscribing legitimacy via the "negotiated dominance of certain flows over other flows"[76] on and across networks. Following Alexander Galloway's assertion that "the limits of a protocological system and the limits of *possibility* within that system are synonymous,"[77] I contend that protocological media operationalize the classificatory and separative regimes of counterinsurgent visuality, creating 'legitimacy' as the effect of a homeostasis sustained through the technical surveillance and management of insurgent countervisualization. However, driven by the forces of globalization, growing dissent, and the decentralizing effects of social media, the decreased ability to suppress insurgency and maintain this homeostasis has led to what Mirzoeff, writing in 2011, called the developing "crisis" of visuality. As a technocratic management style best suited to chaos, counterinsurgency re-situates classification and separation as both the *means* and *ends* of visuality, failing to aestheticize a social and political reality in legitimation of its authority to the point where its

* In US military doctrine, this legitimacy stems from the winning of "hearts and minds," meaning the production of a host-culture that is amenable to neoliberal governance.

*un*reality ultimately becomes apparent—creating opportunity for its countervisualization. Absent the palliative of homeostasis, the visuality of counterinsurgency thus requires chaos as the sole condition for its perpetual re-authorization, finding recourse only in the appropriation of insurgency as the discursive justification for a naturalization of the "disequilibrium of forces manifested in war."[78]

Acting as something of a bridge between part one's discussion of visuality and media and part three's account of the intensification of COIN visuality through the regulatory, technical, and epistemic closure of generative AI, part two of the dissertation correlates the crisis of visuality described by Mirzoeff to the ongoing crisis of establishment legitimacy. The section opens with the contention that the crises of visuality and neoliberalism are essentially interchangeable insofar as both stem from the increasing inability of an established order (both discursive and political economic) to manage the social and legitimate its authority, creating a desire for restoration pursued through intensifications of counterinsurgent control. Characterized by widespread disillusionment with establishment narratives, the crisis manifests as the breakdown of discursive control and a rise in insurgency as changing media conditions make it possible for previously-atomized individuals to form politically salient *aggregates*, leading to the global emergence of (largely right-leaning) anti-establishment movements which in turn threaten the position of incumbent political regimes.

Countering mainstream narratives of the crisis (whereby a credulous public has been 'misled' by the unchecked spread of 'disinformation') the section presents an argument that platforms, in their function as "homophilic clustering technologies,"[79] precipitate the crisis not through the spread of false information, but the algorithmic identification and aggregation of subjects around shared discontents with the reality of their political and economic

marginalization. While various political movements then capitalize on the opportunity presented by the grouping of disaffected publics into demographic packages ripe for operationalization, the reactionary disinformation narrative functions to de-politicize dissent and lay a discursive groundwork for the securitization of digital communications, declaring the *existential threat* of a 'disordered' information ecosystem in justification of extra-political intensifications of counterinsurgency in the form of new information controls.

Applying the concept of securitization,* part two interprets the existential rhetoric around generative AI as a series of 'moves' by which establishment actors are able to discursively transform the governance of new information technologies into a matter of security requiring swift and decisive action. When successful, processes of securitization authorize the introduction of countermeasures which circumvent politics, unilaterally reallocate public resources, and/or infringe on otherwise inviolable rights.[80] Following this, the claim of generative AI's securitization is qualified through reference to research demonstrating high levels of public concern about the technology, as well as a marked increase in support for the imposition of new controls "even if it limits people from freely publishing or accessing information."[81] Given a lack of empirical evidence to support the narrative that 'uncontrolled' generative AI meaningfully increase the supply[82] of and demand[83] for disinformation, such establishment histrionics might then be described as the invocation of a new *moral panic* declaring generative AI as one of many "folk devils" which threaten the social order.† As the only recourse for a beleaguered establishment committed to the technocratic management of an increasingly

* In the context of this dissertation, securitization refers to the term as it is defined in the fields of international relations and security studies, which bears no relation to those more commonly associated with finance.

† See the literature review for a more full accounting of this framing.

unpopular and dysfunctional complex, the inducement of moral panic thus creates the pretext for an intensification of counterinsurgent control through the deployment of the 'safe' generative model as a means of discourse regulation. As both a discursive and technical device, I contend that the development and deployment of the safe generative model is a central to the establishment goal of returning the public arena to homeostasis,* threatening the emergence of a communications infrastructure which performs counterinsurgency as a matter of course.

Finally, having presented crisis as the political and discursive grounds for intensification, part three of the dissertation applies the analytical framework developed in part one to an examination of generative AI through the lens of counterinsurgent visualization. This analysis begins with a reminder that the technologies we call generative AI belong to a specific subset of machine learning and therefore share those technology's inherent predisposition toward the reproduction and amplification of whatever statistically dominant patterns are observable in their training data. While this characteristic alone reveals a relation between generative media and dominant regimes of visualization, I go further to suggest that generative AI differs from prior media in that it is uniquely capable of functioning as a technical agent of visuality's intensification (again, in the Foucauldian sense of increased efficiency and saturation—a mutation of power such that it increases its effects "while diminishing its economic [and] political cost").[84] Whereas the technologies of inscriptive and indexical media were *complementary* to the exercise and legitimation of plantation and imperial authority, I contend that generative models represent a more literal *materialization* of visuality, intensifying the

* Characterizing the establishment's counter-disinformation effort as a "war of restoration," Jack Bratich describes this goal as the "[re]unification of […] society around a political spectrum dominated by a center."; Bratich, "Civil Society Must Be Defended […]": 312.

discursive complex of counterinsurgency through the operationalization of a parallel technical regime of classification, separation, and aestheticization.

The section then points to bullish corporate narratives and growing product integrations as evidence of the model's technical and discursive positioning as a ubiquitous feature of future digital infrastructures. As a result of their conception and development as intermediary platforms, I argue that the generative model is well-placed to naturalize and extend control, making it possible to address insurgency through the precise application of force while still presenting as a smooth and permissive interface. After a brief elaboration on counterinsurgency vis-a-vis the crises of legitimacy and insurgent countervisualization, the strategy of *disaggregation* is presented as the structuring framework for an analysis of generative AI's strategic deployment in the contexts of disinformation and political instability. Understanding insurgency as a disordered 'system state' where preexisting elements in a society "organize themselves in new patterns of interaction," strategies of disaggregation seek the restoration of order through tactical disruption of the insurgent's ability to aggregate and function as a system.[85] Of the various modes of 'attack' available to the counterinsurgent, disaggregation emphasizes the 'choking off' of inputs (in the form of money, personnel, and information); interdiction of links (disrupting lines of communication between individual insurgents and insurgencies); denial of outputs (curtailing the spread of insurgent narratives); and the mobilization of a "single narrative" which excludes the insurgent.[86] Using these methods as a framework, the remainder of part 3 presents an overview and analysis of the regulatory, technical, and epistemic closure of generative AI as co-constituting a wider complex of counterinsurgent disaggregation.

Organized around the principles of 'safe and responsible' AI, regulatory controls on generative AI function to criminalize, discourage, or otherwise impede the development and/or deployment of high-performing models that do not conform to (often broadly defined) safety standards. As such, new and proposed regulatory frameworks from the European Union, People's Republic of China, and the United States effectively "choke off" insurgent visualization through the creation of mechanisms for compelling AI developers and service-providers into compliance. Through various training data restrictions, reporting requirements, contract denials, and an ever-expanding body of trade controls, regulators are able to direct resources away from non-compliant developers of "unsafe" generative models and thus secure the dominance and influence of well-established companies with shared political economic interests. As such, regulatory controls are intimately related to the technical closures which ultimately 'produce' the safe model.

In order to be brought into compliance (or *alignment*) with regulatory and philosophical standards for safety, developers of generative models have devised an array of weight-level (data and training augmentation) and inference-level (intervention at the point of generation) controls. Operating on the principle that generative AI cannot mimic patterns it hasn't observed, data-based approaches involve the removal of problematic, low quality, or otherwise undesirable information from the data used to train a model—as seen in the open-source 'Colossal Clean Crawled Corpus' and its guiding "List of Dirty, Naughty, Obscene and Otherwise Bad Words,"[87] or the Chinese government's "corpus source blacklist" outlawing the use of "illegal and unhealthy information" to train domestic AI.[88] In addition to data-filtering, training-based methods like Reinforcement Learning from Human Feedback 'fine tune' the weights of pre-trained models by assigning a reward function to 'aligned' outputs as determined by carefully-

selected human labelers. On the other hand, inference-based approaches to alignment include real-time moderation of model-user interaction by sidecar classification models, shaping model behaviour through a hidden set of instructions called the *system prompt*, and invisibly modulating or otherwise altering the form and content of user requests before they're submitted to the model for generation. As such, aligned models enable disaggregation through the interdiction and denial of the communicative links and outputs that sustain insurgency, limit the capacity for countervisualization, and exhibit counterinsurgent control as an emergent property.

Operating alongside the aforementioned regulatory and technical countermeasures, epistemic closures support disaggregation through the delegitimation and exclusion of insurgent visualization, effecting the mobilization and technical enforcement of a "single narrative" for interpreting and assigning epistemic value under conditions of contested legitimacy. At their core, epistemic measures influence user credibility assessments through the creation of new techniques and standards for media provenance, authenticity and labeling. At the time of writing, the predominant framework for tracking and attributing provenance to digital assets is the technical standard developed by the multi-stakeholder Coalition for Content Provenance and Authenticity—or C2PA.[89] As a protocol for "storing and accessing cryptographically verifiable information whose trustworthiness can be assessed based on a defined trust model," the C2PA specification provides a technical framework for creating a manifest (also called a "Content Credential") made up of statements (or "assertions") about the origin and history of a given digital asset—including declarations about the history and source of AI-generated media. In light of criticisms of its security and actual ability to cryptographically verify data, the specification's capacity to function as a means of epistemic regulation is revealed as being reliant on a conflation of *provenance* credibility with the orthogonal concept of *content* credibility, creating a

sociotechnical protocol which bestows "certified" media an implied truth affect while simultaneously casting doubt on the authenticity and value of visualizations from outside its regime of legitimacy.

Part three of the dissertation continues with a brief discussion of the Christchurch Call,[90] Tech Against Terrorism's 'Terror Content Analytics Platform,'[91] and the US Special Operations Command's commissioning of an "Automated Orchestration Platform"[92] as early examples of generative AI's unambiguous operationalization in service of the wider counterinsurgent effort to tackle disinformation, radicalization, and the spread of 'terrorist and violent extremist content' via online platforms. To the technophilic OSINT imaginary, generative AI facilitates disaggregation by not only allowing security actors to identify and block the upload and dissemination of certain content, but in its visualization of the network as a "fully rendered actionable space"[93] within which it becomes possible to disrupt insurgent patterns of interaction, return homeostasis, and 'hold' the resulting regime of legitimacy through techniques of modulatory control. Following this, I conclude that 'safe' generative models—achieved through closure—both produce, and are the product of, the reactionary intensification of counterinsurgent visualization by a political economic establishment in crisis. Grounded in the denial of the right to look—understood as the claim to "a subjectivity that has the autonomy to arrange the relations of the visible and the sayable"[94]—closure stems from authority's imagined equivalence between the exercise of this autonomy and the catastrophe of infocalypse. Central to an emergent infrastructure offering cultural-production-as-a-service, safe models ensconce disaggregative control in the technical substructure of knowledge and communication, render counterinsurgency "a regular function, coextensive with society,"[95] and shore up establishment visualization against mounting dissent.

Following a coda using visuality as a lens for analyzing the increasing importance of generative AI in geopolitical and ideological contest, the dissertation conclusion presents the speculative image of a fully realized COIN infrastructure with the 'safe' generative model at its core. Through reference to Nicholas Negroponte's vision of the *interface agent* and Neal Stephenson's 1995 novel *The Diamond Age,* I illustrate the generative models' pervasive surveillance and modulation of communications and subjectivity as automating the management of dissent and inducing in the public a state of political *aphantasia*.[*] Disaggregated and dispossessed of the right to look, the aphantasic body politic is one incapable of visualizing a world apart from the existing realities of establishment visuality, which in turn draws authority from its management of the otherwise heterogeneous terrain of culture.

As a counter to the seemingly inexorable intensification of counterinsurgency, the dissertation concludes with a gesture to otherwise possibility. As much as generative models diagram and extend the classificatory and separative regimes of counterinsurgent visuality, they simultaneously reveal this visualization as the product of an ultimately *mutable* technical system. Whereas counterinsurgency imagines it as control, projects of countervisualization might appropriate the model as a technology capable of empowering the right to look, collectivizing data and claiming the authority to visualize and perform new social and political realities. As such, future work on the subject might focus on developing a theory and practice of generative AI as fundamentally speculative technologies, countering a reality that "did exist but should not

[*] First described by Francis Galton, aphantasia is a cognitive condition characterized by the inability to visualize mental images. Individuals with aphantasia are unable to create pictures in their mind's eye, despite being able to think and reason normally.

have”[96] and reconfiguring visuality toward the visualization of realities that should exist but are as yet becoming.

LITERATURE REVIEW

Introduction

As a study of generative AI's development and deployment as a political technology, this dissertation is located at an interstice of disciplinary and thematic terrains and thus benefits from and seeks to extend a body of knowledge comprising the efforts of thinkers working across a number of related fields. As a political economy of media, it builds on historical and contemporary examinations of how power operates on and through technologies and practices of communication to reflect, reinforce, or challenge the distribution of political, economic, and discursive power in technologically-advanced societies. As a work about artificial intelligence, this project is also in conversation with the ideas of scholars involved in the still-developing field of critical AI, whose humanistic and technical critiques of AI systems and the sociopolitical, economic, epistemic, and environmental consequences of their deployment provides a solid foundation from which to further a critique of the technical and political operations of 'safe' generative models. Finally, by situating this critique in the thematic context of disinformation and the crisis of establishment legitimacy, the study also draws on and contributes to a growing body of critical research questioning the assumptions underpinning orthodox definitions and explanations of disinformation as a political, epistemic, and technical phenomenon. With this dissertation, I intend to contribute to the (rapidly) growing number of contemporary studies on generative AI. Combining the aforementioned disciplinary and thematic influences to broaden

understanding of the model's development and deployment as a political technology, it utilizes the theoretical frameworks of visuality and counterinsurgency as devices for tracing an under-articulated, but operative, thread across political economy, critical AI, and critical disinformation.

Political Economy of Media

Political Economy

Political economy can be a somewhat ambiguous framework of analysis, laden with different meanings and implications according to the specific disciplinary contexts in which it's applied. Rooted in classical economics, it focuses on examining the relationship between politics (understood as the negotiation, distribution and exercise of authority and decision-making) and economy (the production and distribution of goods and services) as a reciprocal dynamic affecting the organization and allocation of power, resources, and value within market societies.[1] Applied to the study of media and communications, the framework analyzes how political and economic power relations—expressed in conditions around media ownership and control, profit incentives, regulation, reach, and consumption patterns—influence public discourse and shape the production, distribution, and reception of information in media-rich environments.[2] In assembling the texts for this section, I have selected a handful of foundational and particularly salient contributions to the political economy of media that I believe efficiently scaffold the later review of critical AI. Given the subject and scope of this project, my aim here is not to be exhaustive but to construct a conceptual foundation for the more specific political economic

concerns of that field while prioritizing frameworks pertinent to the regulatory, technical, and epistemic controls taken up in part three of this dissertation. Despite disciplinary variances in emphasis and approach, political economic analysis can be summed up as the study of what and whose interests and/or goals are considered or prioritized in the allocation of supposedly limited societal resources—material, informational, or otherwise. In each case, the negotiated distribution of scarce resources is thus understood to have political and social consequences unaccounted for by strictly economic analyses of their exchange.

Historically, mainstream political economic analyses of media and communications can be broken down into two strands: 'traditional' and 'critical.' According to traditional interpretations, neoliberal tendencies toward the deregulation and commercialization of communication are said to cultivate an independent media that, by virtue of its economic reliance on advertising alone, evolves to be both representative of and responsive to the values and concerns of its public.[3] Critical perspectives challenge this interpretation, arguing instead that deregulatory market conditions and industrialization of the press have created a media dependent on the support of dominant political and economic interests.[4] It is claimed that, due to the increased costs and concentration of resources required for publishing, journalism and media companies thus evolve to align with government and corporate interests in order to secure the funding, contracts, and access necessary to compete in the market and generate advertising revenue in the first place. This 'free market' analysis of media influence was perhaps most famously articulated by Edward S. Herman and Noam Chomsky, whose 'propaganda model' identified five key "filters" by which the mass media is shaped to reflect and support the interests of dominant state and private entities. These include (1) concentrated ownership and profit orientation within dominant media conglomerates; (2) the reliance on advertising as the primary

revenue source; (3) systemic dependence on information provided by governments, corporations, and credentialed experts aligned with power; (4) the disciplining effect of "flak," or negative responses from influential actors aimed at policing content; and (5) the mobilization of ideological threats—originally framed as "anticommunism"—to set the bounds of acceptable discourse.[5]

Despite being formulated in the context of late-twentieth century mass media in the United States, the model of influence described by Herman and Chomsky remains relevant to analyses of communications well into the twenty-first, with perhaps only "anticommunism" requiring update as the 'existential threat' du jour. As an early and cogent account of media influence, *Manufacturing Consent* is a central pillar in the tradition of critical political economy (CPEC) to which this project both belongs and is indebted.

Undoubtedly, critical frameworks first developed in order to scrutinize the dispositions of mass media in the mid-to-late twentieth century remain salient to analyses of how external political and economic pressures continue to shape communication despite trends toward decentralization and other radical shifts in media technology and practice. Writing on the impact of the internet and early social media on political organization in the United States, David Karpf observed in 2012 that even under changing technical conditions the political *outcomes* of communications remain subject to established structural constraints, noting that "new channels for political speech fit into an existing system of powerful actors and institutions" who in turn influence the shape and character of communications and technology.[6] By this implication, the core concerns of 'traditional' CPEC remain valuable to the macro study of technologies which bear little or no resemblance to those of the mass media era. However, as we shift focus toward digital information ecosystems, including the role of algorithms, platforms, and generative

artificial intelligence, these core principles fall short of fully accounting for the distinctive characteristics of new media—namely, their decentralized nature, automated decision-making processes, and the disintermediation of traditional media structures.

Writing in 2014, when the implications of algorithmic media were only beginning to become more widely understood, Tarleton Gillespie provides a rationale for the political economic interrogation of algorithms as media, having become a central feature of digital information ecosystems. He argues that algorithms should be understood as a *communication technology*—on par with broadcasting and publishing—rather than merely technical systems. As such, they function as what Gillespie, borrowing from earlier media theorists, calls "the scientific instruments of a society at large," embedded within and shaping the social ratification of knowledge.[7] Set apart from their study as purely technical phenomena, Gillespie argued for a "sociological analysis" of algorithmic media oriented toward unpacking the "warm human and institutional choices" which influence their supposedly 'objective' mediation of information and communication online.[8] As a work of political economy, such an analysis would thus examine the 'choices' or objectives behind the development and deployment of algorithms in order to understand their effects on communication, information processing, and sociality in relation to external considerations of ownership, profitability, regulation, and control.

While not explicitly about algorithms, Yochai Benkler's *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (2006) offered a contribution to such a political economy of networked media, arguing that early technologies of decentralization had the potential to shift power away from established centers by enabling a model of what he called 'commons-based peer production'. This model presented a mode of production which, owing to the decentralized nature of networked communication, was built on the voluntary work of

distributed individuals now able to organize production without the coordinating effects of the market or managerial control.[9] While Benkler's optimistic account of the internet as a digital commons presaged a restructuring of the rights to access, use, and control informational and communicative resources, it was criticized for a lack of attention to the ultimately proprietary nature of the internet's underlying infrastructures, enabling the misguided assumption that decentralized networks invariably tend toward non-hierarchical organization. In one such critique, David M. Berry suggested that Benkler's over-emphasis on the radical potentiality of peer-production came at the expense of recognizing that networks, if so highly-productive, are likely to be absorbed into the already existing hierarchies of industrialized production—punctuating the point with a wry paraphrase of Steve Jobs: "the corporate world may soon provide peer-production for the rest of us."[10] As we've seen, Berry's foreboding critique has proven prescient, and many political economic analyses of the platform internet have since traced the state/corporate appropriation of networked peer-production as means of generating wealth and consolidating control.

For example, Christian Fuchs' *Social Media: A Critical Introduction* (2013) provides a Marxist analysis of how algorithms work to stratify the internet as a sphere of production and thus facilitate the collection, commodification and exploitation of user labour.[11] Highlighting their co-option as engines of capitalist wealth generation, Fuchs illustrated how the algorithmic governance of early social media and content-sharing platforms was predicated on the surveillance and expropriation of user data, and thus comprised a networked apparatus which simultaneously reinforced political and economic domination and limited the democratic potential of decentralization.[12] More recently, Shoshana Zuboff has further examined the role algorithms play in the broader political economy of digital capitalism, claiming that the social

and economic pre-eminence of the platform model has given rise to and intensified an entirely new form of capitalist production. Facilitated by post-9/11 policy changes which enabled the relatively unchecked expansion of the internet as a surveillance apparatus, Zuboff contends that tech companies like Google and Facebook have constructed an economic model around the packaging and sale of user data as "prediction products" for use in advertising and other "third-order operations of modification, influence, and control."[13] Shifting the locus of power from ownership of the means of production to algorithms as the means of behaviour modification, Zuboff constructs a political economy of the modern internet characterized by the algorithmic production and exercise of an "instrumentarian power" which surveils, predicts, and modifies user behaviours as though they are technologies themselves.[14]

I present the above as rooted in more 'traditional' CPEC analyses of networked media insofar as their explanatory frameworks largely refer to modes of economic and political influence that were also operative in the study of pre-digital communications. With regards to algorithms specifically, the insights of Fuchs, Zuboff, and others effectively illustrate how the 'filters' of concentrated ownership, economic reliance on advertising, and overlapping state/corporate interests continue to shape the political economy of media and communication under conditions of decentralization. Prioritizing engagement, profit generation, and control over and above the ideal of communication as a public service, algorithmic media thus become 'responsive' to their publics only insofar as increased personalization serves those objectives. But, as mentioned earlier, the core principles of CPEC fall short of fully accounting for the distinctive characteristics of algorithmic media—a shortcoming only further exacerbated by the emergence and ubiquitous integration of more complex machine learning technologies into information ecosystems. This is not intended as a critique of the previously mentioned studies

but rather serves to locate them at a point of overlap or transition between CPEC analysis and more medium-specific approaches to understanding the political economy of AI.

Inherently Political Technologies

As algorithms proliferate in the technical substrates of modern sociality, understanding contemporary political economy increasingly requires a concerted engagement with the functional and communicative role of code and the knock-on effects of subjecting human discourse and knowledge to the procedural logics of computation.[15] In continuity with his earlier critique of Benkler's inattention to the materiality of networks, Berry emphasizes the need to interrogate not just the sociopolitical contexts surrounding the development and implementation of digital technologies but their *computationality*, that is, to attend to the technical structures and processes of software as determinants in and of themselves.[16] Such an analysis thus seeks to make legible the otherwise inscrutable politics embedded in code, revealing how the operational logics of technical systems function to shape outcomes, sometimes independent of explicit intention or design.

In his influential essay "Do Artifacts Have Politics?" Langdon Winner answers his own rhetorical question, stating that "the things we call 'technologies' are ways of building order in our world," and therefore inescapably imbricated with the political.[17] In illustration, Winner provides the example of Robert Moses' construction of bridges along the Long Island Parkway which, due to their unusually low clearance, made the newly-founded Jones Beach State Park

(1929) virtually inaccessible by bus.[18] As a result, while the white middle-classes were able to use the parkway for commuting or recreation by car, poor and racialized citizens (who disproportionately relied on public transit) were effectively barred from accessing the area. Winner argues that, by virtue of their specification, the parkway overpasses functioned to *encode* the politics of inequality into the materiality of urban space, deploying public works as "a way of engineering relationships among people that, after a time, becomes just another part of the landscape."[19] According to Winner, there are two ways that technologies come to have political qualities. Owing to evidence that Moses was deeply prejudiced along lines of race and class, he presents the parkway bridges as an example of a device or system *intentionally* designed to impose a particular social or political order. In contrast, Winner also introduces the idea that some technologies are *inherently* political—meaning they both require and reinforce particular political and social relations as essential conditions of their operation.[20] Through Engels, Winner offers the industrial-era factory as an example of a technology whose imposition of a social order is not accidentally or situationally determined but is rather its *condition of possibility*. Accordingly, the political economic analysis of code, platform algorithms, and machine learning technologies like generative AI require an attentiveness to both external and internal conditions.

Despite being radically different in terms of their materiality, the parkway bridges and computer code are ultimately quite similar in how they circumscribe and enforce their politics. As the short bridges govern access and movement according to an externally imposed system of classification, codes similarly control flows of information according to explicitly programmed objectives and the Boolean logic governing which connections are permitted and which are restricted. This similarity prompted Lawrence Lessig to make the argument that code is the "newly salient regulator" of human activity, announcing: "'Lex Informatica,' as Joel Reidenberg

first put it, or better, 'code is law'."[21] Like law in the legal sense, code can be *intentionally* political, in that it functions to regulate activity in ways that reflect the values of its authors. While, in response to criticism from legal theorists, Lessig asserts the productivity of temporarily "ignoring" differences between code and law, Wendy Chun contends that code's effectiveness as a regulator actually stems from a fundamental dissimilarity. Casting code as both judge and constable, Chun emphasizes code's unique ability to enforce *itself*, encoding law as "an inhumanly perfect 'performative' uttered by no one."[22] This self-enforcing nature of code, she argues, highlights how its power lies not only in its capacity as *intentional* regulation but in its subsequent autonomy and ability to eschew human accountability, creating systems where influence can be subtly but firmly exerted without the direct intervention of its creators.

In *Protocol: How Control Exists after Decentralization* Alexander Galloway constructs a political economy of the internet through a technical and theoretical analysis of the codes, or protocols, which construct it as a stable, navigable network of connections between computers. As Galloway explains, connectivity on the so-called "clear" internet* is governed according to a negotiated balance between two contradictory protocols: TCP/IP (Transmission Control Protocol/Internet Protocol), which enables unencumbered peer-to-peer connection across a horizontal network; and the DNS (Domain Name System), which controls this connectivity by mapping the network as an inverted-tree wherein each node is assigned an address.[23] While the DNS protocol is critical for enabling individual nodes on the network locate one another and exchange information, Galloway emphasizes that its hierarchical structure means that control is consolidated in nodes the further they are up tree. As the set of rules for ordering the otherwise

* As opposed to the "deep" web, comprised of websites not indexed by the Domain Name System.

anarchic interactivity enabled by TCP/IP, the DNS protocol is thus imposed as a "distributed management system that allows control to exist within a heterogeneous material milieu."[24] Thinking back to Winner's distinction, the stratification of networked connections via protocols like the DNS bestow the modern internet an *inherently* political quality, insofar as its adoption "actually *requires* the creation and maintenance of a particular set of social conditions as the operating environment of that system."[25] As the possibility of connection between client and server is sustained protocologically, the secondary experience of unrestricted flow exists only as an aesthetic effect, produced through adherence to the strictly-defined rules which enable those properties to emerge.

Protocol and code are not strictly interchangeable. While protocol is about the broader, structuring rules for interaction between systems or users on a network or platform, code refers to the specific, localized instructions that allow those systems to operate internally. Despite this, both protocol and code can be mapped as a political metaphor for the regulation of actions and behaviours by or within a given system or society, circumscribing what can or cannot be done, reinforcing certain behaviors to the exclusion of others. As Winner reminds us, while the consequences of some devices (like bridges) vary to reflect the situation and intentionality of their deployment, *inherently* political technologies don't allow for such flexibility, and that "to choose them is to choose a particular form of political life."[26] As complex systems manifested through code and governed by arrays of training, interaction, and protocol, the political economic analysis of generative models requires attention to both their intentional and inherent politics; regulating knowledge, communication, and action as the combined effect of their deployment in specific political contexts to specific political ends, and an inherent tendency to *visualize* according to particular patterns of relation. This project thus seeks to map the external and

internal conditions of generative AI's deployment in the political context of disinformation and epistemic disorder, characterizing the 'safe' generative model as a technology of counterinsurgent visualization.

Critical AI Studies

Introduction

While 'traditional' approaches to CPEC can provide valuable insights into the political economy of media and communications, the existing frameworks of analysis can fall short in accounting for the complex sociotechnical entanglements of contemporary media—particularly as decisions about the allocation of resources are increasingly automated by machine learning technologies. As such, a growing body of scholarship commonly referred to as 'critical AI' has emerged to address this gap, articulating a political economy that, by virtue of its attention to the medium-specificities of AI, is better attuned to the complex imbrications of external social forces and *internal* technical dynamics which shape them as political technologies.

Critical AI is not a monolithic field. Though its disciplinary boundaries remain blurred, it can be said to encompass two distinct, though interconnected, threads: political economies of AI and philosophies of AI. As already discussed above, the former focuses on the internal and external relations of material, economic, and political power which shape the development,

deployment, and sociopolitical effects of AI systems. On the other hand, philosophies of AI focus more specifically on onto-epistemological concerns, like the question and nature of machine intelligence, shifting relationships and distinctions between human and machine, and the role of AI in knowledge production. While this review will cover some critical perspectives on the onto-epistemological implications* of generative AI, in keeping with the disciplinary scope and register of this dissertation, it otherwise favours works which more concretely interrogate AI as a complex of inherently political artifacts and practices. As such, the following review is structured around several key concepts or themes that I believe best account for the current understanding of generative AI's political economy. These of course include concerns around the political and epistemic effects of rendering the world legible to "machine learners";[27] AI's inherent bias and tendency toward the reproduction and entrenchment of statistically-dominant patterns; the expropriation and control of public data as resource for visualization; and the precautionary logics shaping AI as a means of risk management exercised through surveillance, personalization, and modulatory control. Throughout, I attempt to qualify the dissertation's presentation of *visuality* as a unifying theoretical framework by relating the concepts of visualization, Visuality 1, the right to look and counterinsurgency to knowledge production, bias, data and control, respectively.

Defining Artificial Intelligence

* Insofar as they effect knowledge production and that their world-making capacities can be understood as an "ontological politics," whereby generative systems construct reality through governing the fields of action, possibility, and relation: Annemarie Mol, "Ontological Politics. A Word and Some Questions," *The Sociological Review* 47, no. 1 (May 1, 1999): 74–89.

Despite the increasing ubiquity of the term, and not for lack of effort,[28] there remains little consensus around the definition of artificial intelligence and what specific attributes distinguish it from other modes of statistical or probabilistic computation. This is, of course, partly due to difficulties in developing a workable model of intelligence or cognition in-and-of itself, thus making the secondary task of distinguishing machine intelligence all the more complicated.[29] While lack of an all-encompassing definition might not significantly impede the pursuit of more discrete objectives within the realm of technical research, lack of clarity around what does and does not constitute AI stands in the way of effective collaboration and coordination in advancing understanding of the field as a whole.[30] This is equally important to critical AI scholarship, whose very existence as a discipline implies a degree of consensus around what, precisely, is the object of study. Another side-effect of this imprecision in definition is that—despite its rootedness in a diverse history of technical research—the cultural imagination of artificial intelligence as represented in news headlines, marketing, and popular culture, almost universally "falsely implies something singular and unprecedented."[31]

For our purposes, a working definition of "artificial intelligence" might eschew reference to specific technical or human-like characteristics in favour of a model evaluating systems according to a pre-defined standard of rationality. According to this framework, AI refers to the construction, through varying means, of non-human *rational agents*.[32] Insofar as an 'agent' is anything that *acts*, 'rational' agents are those which act "so as to achieve the best outcome or, when there is uncertainty, the best *expected* outcome."[33] Thought this way, artificially intelligent systems are those that do "the right thing," here understood as the most desirable input/output relation with respect to a specific performance measure or goal.[34] In developing an

understanding of how programs come to function as rational agents, we must also acknowledge a critical distinction between what are known as symbolic and connectionist paradigms of AI.

In symbolic AI, often referred to as "good old-fashioned AI,"[35] rationality derives from a hard-coded system of corresponding symbols and rules, to which it applies deductive *if/then* logic to make inferences and solve problems.[36] While these systems excel in structured scenarios, their reliance on pre-programmed rules limits their ability to adapt to unexpected inputs. In contrast, connectionist AI forgoes the explicit programming of logic, instead constructing its rationality as an emergent property of its encounter with data.[37] Using inductive reasoning, connectionist systems, such as generative models, are thus able to adjust to changing conditions by generalizing from the patterns and associations it observes.

This divergent orientation to rationality has implications for the analysis of AI as goal-oriented systems. In the symbolic paradigm, a system's governing rationality is largely transparent, as its operating logic is directly derived from the relationship between explicitly defined goals and rules. On the other hand, the reasoning of connectionist systems like generative models is generally more opaque, as they leverage statistical inference to continuously develop and adjust their governing rationales through the complex and often inscrutable process of 'learning' from data. While technical histories[38] often portray rational agents as neutral arbiters, critical perspectives on AI contest such claims, highlighting the economies of influence and effect which underlie their real-world deployment. In particular, critical scholars emphasize the sociopolitical contexts of autonomous decision-making, where the divergent and often conflicting interests of parties mean that the actions or policies derived from such systems almost invariably create "winners and losers." [39]

While their modeling and implementation of rational "behaviour" differs from more straightforward symbolic AIs, generative models remain fundamentally 'goal-oriented' systems insofar as the training process autonomously maximizes certain reward functions or objectives—i.e. minimizing error and producing 'realistic' or otherwise 'desirable' outputs.* Because connectionist systems do not operate using symbols or explicitly defined rules, the inherent political character of their generativity is often obscured and/or naturalized as the objective operations of a well-functioning system, thus illustrating the need for medium-specific analyses. A critical first step in knowing how to build, govern, or otherwise live with AI, then, is to develop means to "know them in the first place."[40]

Inscrutability

Perhaps the most influential and evocative metaphor used to describe the inscrutability of technical systems is that of the black box. In *Pandora's Hope: Essays on the Reality of Science Studies*, Latour describes the expression 'blackboxing' as the process whereby "scientific and technical work is made invisible by its own success."[41] What's meant is, when a mechanism or technical artifact works as it should, to use it one need only focus "on its inputs and outputs" and ignore the internal complexity of its operation. For Latour, the effect persists so long as the system operates as expected, and it is only in moments of malfunction that its constitution via the

* For example, in the GAN framework pioneered by Goodfellow et al., the model's ability to generate realistic outputs emerges through a form of unsupervised learning, where the discriminator provides feedback to the generator based on the difference between real and generated data. This feedback then guides the generator in refining its outputs. Ian J. Goodfellow et al., "Generative Adversarial Networks" (arXiv, June 10, 2014).

"joint production of actors and artifacts" is made visible, as a distributed network of heretofore "silent entities" appear to restore it.[42]

While Latour's formulation explains the cultivated invisibility of the black box as a property of the stable artifact, the 'black box' also functions as a technical and discursive device for obscuring the social and political implications of its design and use. By concealing the complexities of how and why a technology functions, the black box contributes to the illusion of neutrality and objectivity, often masking the value-laden decisions and power dynamics embedded within the system. Frank Pasquale suggests that the political nature of obfuscation thus reflects a "general truth" of the technologically-advanced societies we live in today: "What we do and don't know about the social (as opposed to the natural) world is not inherent in its nature, but is itself a function of social constructs."[43] According to Pasquale, these social constructs set technical, legal, and conventional limits on what can or cannot be 'seen' or investigated, constituting an 'asymmetry of knowledge' wherein political and economic power is commensurate with the right to "scrutinize others while avoiding scrutiny oneself."[44] Contrasting the mass collection and analysis of information by 'Big Data' with the relative inability of individuals to assess how their data is used, Pasquale likens what he calls the 'black box society' to a one-way mirror, making it increasingly difficult for citizens to discern the rules that govern their lives, and whose interests they might serve.[45]

While nearly all modern computational systems pose challenges to interpretability, generative AI uniquely presents as the quintessential black box, with their neural networks being "possibly the least interpretable" mode of machine learning.[46] Most often encountered through an API interface, generative models accept inputs in the form of text, audio, or visual prompts, and—seemingly by magic—generate realistic or otherwise coherent outputs. While these outputs

often present as meaningful, the actual relationship between the user's input and the generated data is difficult to trace to specific decisions and/or operations performed by the model. The inscrutability of generative models is two-fold: on one hand, the application interface is often constructed so as to intentionally obfuscate the (often proprietary) techniques and content controls shaping model behaviour, protecting trade secrets and constructing the unencumbered user experience,[47] while on the other, the actual workings of connectionist systems are not readily understood, even when models are open-sourced and thus made available for inspection.

Suggesting that challenges to interpretability run deeper than technical opacity alone, Beatrice Fazi has said that attempts to make sense of connectionist systems are further confounded by the fact that adequate descriptions of model behaviour increasingly exceed the capacities of human language or thought.[48] We may refer again to Latour's initial definition of the black box, which suggests that as machine learning techniques become more successful/capable, the inscrutability of their internal workings increases in kind. Describing this process, Fazi contends that today's most advanced models exhibit a degree of "onto-epistemological autonomy," which is to say the way that generative models 'think' has become fundamentally *incommensurable* with the "epistemic landscape of human cognitive representation."[49] This is perhaps most clearly reflected in the frequent admission by computer scientists that even those involved in developing leading generative models can neither understand[50] nor predict their behaviour.[51] This notion of incommensurability[52] and the barrier it poses to notions of human understanding and control of advanced AI has itself given rise to a subfield of research into so-called *explainable AI* (XAI).[53]

However, in *If…Then: Algorithmic Power and Politics* (2018), Taina Bucher complicates the metaphor of the black box, arguing that algorithms are "neither as black nor boxed as they

are sometimes made out to be," and to frame them as such risks overemphasizing "visibility as a conduit for knowledge and control."[54] While they are generally inscrutable when examined at the level of code, the politics of connectionist* AI like generative models are more often constituted via a wider system made up of technical, institutional, and human actants—the network of 'silent entities' implicated in the stable operation of Latour's device.[55] According to Bucher, efforts to address the power and politics of AI might then circumvent the inscrutability of their code (what they *are*) by becoming attentive to what and how they *do*—what she calls their "eventfulness," describing the relational ontology of their 'co-becoming' with other actors.[56] In this reading, while politics might be *inherent* to AI systems, they emerge "not as *what* algorithms do per se but *how* and under what circumstances different aspects of algorithms and the algorithmic are *made* available—or *un*available—to specific actors in particular settings."[57] As such, Bucher contends that to frame AI as an unknowable technical object is to obscure its otherwise knowable political economy. The growing body of scholarship falling under the umbrella of 'critical AI' thus takes up this work of apprehending the social and political entanglements of these "relatively inert, sequential, and/or recurrent structure of matrices and vectors" we colloquially refer to as *artificial intelligence*.[58]

Critical AI

* Any mention of "AI" from this point forward will be in reference to connectionist systems.

In the introduction to his book *Human Compatible: Artificial Intelligence and the Problem of Control* Stuart Russel (who along with Peter Norvig wrote one of the defining surveys of the theory and practice of modern artificial intelligence)[59] presents a technologically determinist perspective on the effects of AI—commonly held by those working in computer sciences—which suggests that solutions to the alignment problem lie in ensuring machines pursue human objectives, and thus advances a framework whereby intelligent systems are understood to be *beneficial* insofar as "*their* actions can be expected to achieve *our* objectives.".[60] However, by misapprehending their fundamental sociotechnicality, understandings of AI which assume a distinction between system and human objectives necessarily set the stage for frustration when system outcomes diverge from the idealized expectations of otherwise well-meaning researchers. In contrast, critical AI contends that to be disappointed with the outcomes of an AI system is, ultimately, to be confronted with the unadulterated objectives of the societies which produce and deploy them.

To this point, Bernhard Rieder reminds us that statistics emerges from the historical attempt to develop a 'science of the state' as a "rational means to govern, manage, and decide,"[61] and as such statistical technologies like machine learning inevitably orient themselves to the world through the coldly instrumental terms which similarly underly neoliberal status quos: cost/benefit, profit maximalization, and control. Crucially, the 'undesirable' classifications, biases, and reproductive tendencies of *inherently* political AIs are as often *intentionally* operationalized in order to produce or sustain certain political economic arrangements—further complicating decontextualized assessments of AI behaviour. This is clearly visible in the changing content policies of major platforms and tech companies following the re-election of Donald Trump in late-2024, as efforts to reduce discrimination and bias in their automated

systems no longer produce the political currency necessary to buy regulatory favour. As such, political economies of AI are continually shaped and reshaped through renegotiations of state and corporate interest.

Critical perspectives on AI shift the emphasis away from *whether* they pursue human objectives toward the political economic question of *which* human objectives are optimized, for *what* purposes, and *who* gets to decide.[62] As such, a central problematic of critical AI, and one this dissertations confronts, is how to work across disciplines, situating AI as a historically contingent "assemblage of technological arrangements and sociotechnical practices" in order to develop new critical frameworks or "countermythologies" for understanding, developing, and living with AI.[63] For example, Matteo Pasquinelli counters essentialist myths of AI's recency, tracing a genealogy of algorithmic thought wherein AI emerges as the latest stage in a long progression of material and socially-situated practices. Understood as one of many standardized 'rituals' for the "division of space, time, labour and social relations,"[64] the algorithm thus evolves from material practice into the more abstract forms we recognize today, all the while losing none of its political character. In a similar vein, Kate Crawford has advocated for a "topographical approach" to mapping the relational ontology of AI's *becoming* as a political technology.[65] Crawford suggests that, by "seeing AI like an atlas," critical studies of AI are able to map its technologies and applications onto various disciplinary 'landscapes,' becoming attentive to particular details and overlaps so as to trace obscured or otherwise under-articulated threads of political economy across them.[66]

The following sections provide brief overviews of select themes, or landscapes, from the critical study of AI and, after a short introduction articulating how the issue has been framed in established critical AI literature, survey how each has been discussed in relation to generative

models and services, specifically. As previously mentioned, each section concludes with a reference to visuality as a unifying theoretical framework for apprehending the political economy of generative AI under conditions of contested legitimacy.

Visualization: The Epistemic and Political Logics of AI Knowledge Production

The increasing automation of knowledge production in technologically-advanced societies by machine learning systems presents a number of challenges to traditional epistemological frameworks. What does it mean for knowledge to be 'generated' by an algorithm, and what epistemic and political consequences arise from the "somewhat unruly generalization" of machine learning as a primary means by which information is produced and operationalized?[67] While a detailed technical explanation of knowledge production in machine learning is well beyond the scope of this dissertation, I will attempt to characterize it enough to contextualize a discussion of these concerns.

Matteo Pasquinelli and Vladan Joler have indicated that AI, as "an instrument of knowledge," is composed of three primary components: the object (the training data), the instrument (the algorithm), and the final representation (the model).[68] Unlike epistemological frameworks which understand knowledge as the ultimately expansive and flexible product of embodied human experience, interpretation, and reason, algorithmic knowledge production is necessarily instrumental and *compressive* in nature, taking place through a 'learning' process whereby the 'world'—represented as a dataset—is statistically analyzed and re-ordered into a

usable (i.e. g*eneralizable*[*]) model of its patterns and correlations. In making data operable, AI systems perform a "prior ordering of data to afford its traversal," iteratively transposing unstructured data into an "increasingly extensive, heavily coordinated space" called the *latent space*.[69] As AI systems 'learn,' they identify "meaningful patterns"[70] in data and encode them as *vectors*—mathematical trajectories that trace specific features and relationships within the model's multidimensional knowledge space.

In generative AI, a trained model thus consists of both this abstract, high-dimensional 'map' of meaningful patterns and a corresponding statistical model that combines and manipulates vectors to generate new data that reflects these learned structures, whether through direct traversals of vectors in latent space[71] or iterative refinement processes like diffusion.[72] What results is a reconfiguration of knowledge as a *statistical* construct, wherein the assessment of meaning and relevance proceeds as a function of pattern recognition and probabilistic inference. As such, a model can only be said to 'know' something only insofar as it has mapped a sufficiently strong correlation between two or more points, thus exhibiting what N. Katherine Hayles has called a "fragility of reference,"[73] describing the tenuous (and ultimately fictional) link between models and the realities they purport to describe.[†] Justin Joque similarly identifies the epistemic consequences of probabilistic knowledge, whereby models autonomously construct concepts from the correlation of discrete data features (edges, color intensities, shapes): "In a world made of data[,] meaning is a matter of algorithmic output: *a probable cat*."[74] By

[*] Generalization refers to the ability of a model to extend its learned correlations to the analysis of data it wasn't trained on. See: Yeounoh Chung et al., "Unknown Examples & Machine Learning Model Generalization" (arXiv, October 11, 2019)

[†] This breakdown is perhaps most visible in the phenomena known as "hallucination," where generative models confidently produce data that diverges from desirable or correct patterns.

producing ‘knowledge’ as a matrix of statistically dominant patterns, the “epistemic machinery”[75] of machine learning thus presents us with “a regime of normative reasoning that, when in the ascendant, takes shape as a powerful governing rationality.”[76] In addition to these and other[77] critiques of AI’s correlationist epistemology, Critical AI scholars have developed a number of positions on the effects of this refashioning, particularly regarding the political logics inherent to probabilistic knowledge production and its technical and discursive naturalization as the dominant way of knowing and acting in the world.

Large generative models like GPT-4o (the back-end of ChatGPT) have been described as a “blurry JPEG of the web”[78]—a rather evocative analogy that helps characterize them as ‘knowers.’ JPEG compression reduces the file size of images by grouping similar data values and approximating them with a single value. For example, if an area of a photograph contains multiple shades of blue that are *alike enough*, the JPEG algorithm stores them as a single averaged value. If that area also included minor variations of colors *other than blue*, they too may be absorbed into the dominant blue. JPEGs are called *lossy* because they compress information by excluding ‘non-essential’ variations while retaining just enough data to reconstruct an acceptable image.

Generative models function in a similar way, compressing data by smoothing over inconsistencies and retaining dominant patterns. Crucially, this selective forgetting is what makes generativity possible, enabling the model to interpolate and recombine rather than simply retrieve data. In machine learning, the extent of a model’s ‘forgetting’ is balanced so as to avoid *overfitting*, where a model simply memorizes data, or *underfitting*, where it oversimplifies patterns.[79] From this balance, generativity emerges as novel means of knowledge production wherein *construction* displaces *description* as “the epistemic tool of choice.”[80] On this point,

Beatrice Fazi highlights how AI models do not describe or interpret the world but actively *reconstruct* it in abstraction, a process which "assumes a life of its own" and thus shapes not just *what is known* but the nature of knowledge itself.[81] Amoore et al. elaborate on the political logics of generative models, suggesting that rather than reducing or distorting "what the real world is," compression actively constructs a "highly generative, productive and derivative space of possibilities" where certain political structures, concepts, and relations are "brought into being in and through the latent space."[82] Generativity thus proceeds according to a political logic where the estimation of underlying distributions does not merely reveal patterns but actively configures the boundaries of what can be known and acted upon.[83]

This invites a critical question: if AI models shape what is knowable and actionable, what are the consequences of falling outside the distributions that define those possibilities? Or, as Kate Crawford has phrased it: "What epistemological violence is necessary to make the world readable to a machine learning system?"[84] Because AI learns through capturing the 'relevant' features of data, the algorithms that determine relevancy necessarily make political choices, optimizing for certain features while ignoring or actively suppressing countless others.[85] In the parlance of this dissertation, generativity *visualizes* and *holds* its regime of legitimacy as a probability distribution. As the trained model positions relevant features within its vector space, where proximity to dominant patterns determines retrievability, features lacking strong association with these patterns become effectively irretrievable, relegated to the interstitial space between vectors. In their paper "Epistemic Injustice in Generative AI," Jackie Kay et al. adopt the term "hermeneutic injustice"[86] to describe the erasure of marginal perspectives that is so often the result of generative models' parsing of collective knowledge.[87] The violence of this exclusion is twofold, operating both as a function of algorithmic compression, and as an inherent

quality of training data that is itself the result of "a complex, biased and mediated process of social production."[88]

Returning to Crawford's atlas analogy, in the attempt "to capture the planet in computationally legible form,"[89] machine learners presuppose a world that has "already become an image,"[90] and therefore position *making the world smaller* as a precursor to the political logic of generativity.* Critical scholarship warns that such reductionism threatens to render the world through an increasingly narrow epistemic frame, a position Mikael Brunila exemplifies in his description of machine learning as a project of "explaining everything in terms of *one* or *very few*" models.[91] In this sense, to construct knowledge as a function of probability is less to create an atlas *of* the world† than it is a drive to be *the* atlas, determining the content and character of a shared knowledge base "while simultaneously denying that this is an inherently political activity."[92] Writing about the influence wielded by developers and regulators of AI systems, literary critic Hannes Bajohr emphasizes that the technical creation/curation of a *world model* is politically commensurate with the formulation of a *social vision*, concluding that "whoever controls [generativity] controls politics."[93]

In politics, as in machine learning, there is a fundamental disjuncture between reality and its imaginary representation. In terms of political epistemology, Walter Lippmann called this representation the "pseudo-environment," a simplified mental image of the world that allows individuals and collectives to act in the face of overwhelming complexity.[94] According to

---

* Although, paradoxically, generativity is simultaneously reliant on variation, as evidenced by the phenomena of *model collapse*, where model's degrade when trained on generated data. Ilia Shumailov et al., "The Curse of Recursion: Training on Generated Data Makes Models Forget" (arXiv, April 14, 2024)

† As imagined by the computer sciences in the form of a *world model* as a kind of internal "simulator" of reality: Yann LeCun, "A Path Towards Autonomous Machine Intelligence," Fundamental AI Research (Meta, June 27, 2022): 7.

Lippmann, this representation is a fiction that both shapes and is shaped through culture, understood as a product of "the selection, the rearrangement, the tracing of patterns upon, and the stylizing of […] 'the random irradiations and resettlements of [knowledge]'."[95] This process is strikingly parallel to compression and vectorization as means of making the world computationally operable. Just as pseudo-environments structure political cognition, vectorized space serves as the epistemic enclosure of AI systems, naturalizing a hegemonic regime of a priori *visualization* circumscribing the limits of knowledge, action, and possibility.

Contra the naive realism of AI evangelists, who assume that algorithmic representations might faithfully capture reality, critical perspectives reveal such models as inherently normative, actively constructing and foreclosing particular epistemic and political formations.[96] Accordingly, AI models are said to perform *interested* readings of reality,[97] visualizing the world from a perspective always reducible to, in Pasquinelli's phrasing, "the eye of the master."[98] As such, Leif Weatherby has called generative models (GPT's, specifically) the "first quantitative producers of ideology" insofar as their vectorial space encodes the hegemony of normative semantics in "pre-digested form."[99] At the same time, Weatherby suggests that generative models also allow the epistemic structure of hegemony to be queried, as the vector space presents an unprecedented opportunity to "generate and then examine 'what is near what' in political semantics"—but only for a limited time.[100]

While existing literature illuminates the epistemological and political ramifications of machine learning's reconfiguration of knowledge as a product of probabilistic reasoning, I contend that an analysis of this 'learning' through the lens of visuality—and so understood as a way of 'looking' that manifests the authority of the visualizer[101]—reveals that the political logics inherent to generativity emerge not at the level of content, but through the intensification of a

mode of *visualizing* legitimacy according to the counterinsurgent imperative to 'clear, hold, and build' order from heterogeneous data. Read this way, generative models intensify and obscure their politics in the incommensurable process of vectorization, a function of the literal and metaphorical distancing of the "place" of visualization from the subject being viewed,[102] and thus shaping generative logics as "a visualized authority whose location not only cannot be determined[,] but may itself be invisible."[103] While Amoore et al. come closest to this formulation (both in the "World Model" paper's account of generativity, and in *Cloud Ethics*, where she locates AI's potential violence in the act of horizontal arraying that enables pre-emptive governance),[104] and Mackenzie and Munster have described the operationalization of visuality by algorithmic assemblages as a new mode of "invisual perception," what they've called "platform seeing,"[105] my project contends that this form of perception is not a novel development so much as a continuation and intensification of counterinsurgent visuality—the technical permutation of an already well-established regime of visualization and control.

I believe that framing AI knowledge production as proceeding from *counterinsurgent visualization* is critical to apprehending not only the inherent politics of generative models, but the political economy of their intentional deployment under conditions of contested legitimacy. Amoore et al.'s analysis of the political logic of generativity allows us to see how Visuality 1 persists independent of its discursive embedding as labels/classifications in training data, but falls short clearly articulating its political stakes. Amoore almost says it, discussing the logic of *generativity* as producing "LLM-derived courses of action [that] are better understood to *be contingent probabilities that nonetheless become political decisions that radically foreclose alternatives,*" or as *sequences* affirming "the capacity to predict the next thing" as a broader logic of ordering the world.[106] This is counterinsurgency, and I contend that naming it as such enables

a more precise theorization of generative AI as apparatuses attuned not only to the logics of prediction and pre-emption, but to the political grammars of legitimacy and stabilization.

Visuality 1: Bias and Discrimination in Generative AI

Being fundamentally probabilistic in nature, generative AI's have famously been dubbed "stochastic parrots," referring to the observation that their outputs tend to reproduce the statistical distributions observable in their training data rather reflect any independently arrived at notion of knowledge or truth.[108] One of the most widely observed *consequences* of this tendency is the persistent amplification of biases and discriminatory outcomes in AI-generated content and decision-making. As critical AI scholars have argued, this bias is not merely a technical issue but an epistemic and political one, actively shaping what is rendered visible, sayable, and legitimate—with tangible social and political effects.

"To classify is human."[109] In their foundational study, *Sorting Things Out: Classification and Its Consequences* (1999), Geoffrey Bowker and Susan Leigh Star acknowledge classification's centrality to the human endeavors of knowing and acting in the world. We sort laundry, organize inboxes, enter washrooms by gender, label and pair objects by specification, etc.; these and countless other performances make up the inescapable and often invisible classificatory work that allows us to make sense, prioritize, and administer the intricacies of individual and collective life. Published at the turn of the millennium, Bowker and Star's study examined the histories and material consequences of medical, scientific and race classification at a time when information systems were undergoing rapid change. Their analysis of how the

formalization of ad hoc classificatory schemes as features of the built information environment "valorizes some point of view and silences another" and thus stabilize social and political categories while disappearing "into infrastructure, into habit, into the taken for granted," is a touchstone for later studies examining bias and discrimination in technical systems.[110] Echoing Bowker and Star's recognition of its ubiquity, Caliskan et al. have noted that, as the (often latent) semantic features of human culture, classificatory biases "should be the *expected result* whenever even an unbiased algorithm is used to derive regularities from any data."[111] In this sense, bias is not so much an aberration as it is an indicator of the system's aptitude at discovering, surfacing, and extending the classificatory schemes embedded in the world as data.

Of course, despite (or rather, *because* of) their ubiquity, classification has wide-ranging social and political effects, many of which are made increasingly visible or otherwise amplified as an "intrinsic discriminatory feature" of AI's development and deployment in the capitalist context.[112] Researchers identified and were studying bias in computer systems as early as 1996, when Friedman and Nissenbaum developed an analytical framework breaking bias down into the interrelated categories of *pre-existing*, *technical*, and *emergent*, allowing bias to be traced from its social roots, through the technical constraints of the system, to its reemergence through user interaction.[113] In more recent years, other foundational research has examined the effects of persistent bias in AI systems, shaping critical debates on fairness, accountability, and linking technical bias to broader histories of oppression.

For example, mathematician Cathy O'Neil's *Weapons of Math Destruction* (2016) and Virginia Eubank's *Automating Inequality* (2018) offer early mainstream illustrations of how the use of opaque predictive models in policy, finance, employment, policing, and healthcare serve to reinforce systemic entrenched inequalities and disproportionately harm poor and working-

class citizens while operating under the guise of neutrality.[114] Both published in 2018, Safiya Noble's *Algorithms of Oppression* and the paper "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" by Buolamwini and Gebru are seminal studies on the racial and gender biases inherent to search engine and facial recognition algorithms and the consequences of their governing the terms of recognition, representation, and discoverability of women and people of colour.[115] In *Race After Technology* (2019), Ruha Benjamin advances the concept of the 'New Jim Code' to describe the employment of "objective" machine learning technologies that sanitize and reproduce the discriminatory practices of a supposedly disavowed racist past, drawing a degree of functional equivalence between the "spurious correlations" which animate corporate algorithms and oppressive practices of race-based segregation.[116] Wendy Chun—also highlighting racist and eugenic histories of classification and correlation—has suggested that reinforced segregation presents as a technical *goal* of AI systems insofar as the "clustering" of people and information is fundamental to prediction and its operationalization as social engineering.[117]

Unsurprisingly, a vast body of research has documented the outputs of generative models as also exhibiting biases across various modalities. In terms of visual generation, racial and gender biases have been observed across model architectures. In the GAN[118] and diffusion[119] frameworks for image generation and, more recently, text-to-video models like OpenAI's Sora,[120] generated outputs have been observed to disproportionately reflect the demographic and cultural biases present in their training data, frequently over-representing whiteness and gender stereotypes while marginalizing or distorting depictions of underrepresented groups. Large language models for text generation have also been noted as exhibiting similar behaviours via the 'embedding' of social and political biases inherent to their training corpora.[121] Additional studies

have examined the propensity of LLMs to 'pick up' selective political biases from "partisan" pretraining corpora, reinforcement learning through human feedback, and secondary classification models for moderating hate-speech and other 'harmful' information.[122] As such, mitigation efforts have themselves revealed bias to be a thorny technical and political issue, as evidenced by the now infamous example where, following efforts to diversify outputs, Google's Gemini AI generated images of Black, Chinese, and Indigenous Nazi soldiers,[123] attracting new criticism that the chatbot was "woke."[124] On the other side of the spectrum there is some concern about dual-use,[125] as the biased tendencies of generative models can be exploited in order to train hyper-partisan models like RightWingGPT,[126] or otherwise modulate model behaviours according to shifting political climates.[127]

Following the understanding that bias results from the model's perpetuation of statistically and discursively-derived correlations, much work has focused on interrogating the inherent politics of multimodal datasets. In their seminal "Stochastic Parrots" paper, Bender et al. undermine notions of the internet as a diverse data-source, citing an array of demographic, geographic, and methodological factors which narrow representation and encode a dominant/hegemonic worldview.[128] Similarly, Kate Crawford and Trevor Paglen have performed an "archeology of datasets" to uncover how mainstream image repositories used for AI training are irrevocably infused with a normative politics through the processes of collection, classification, and labelling,[129] and Miceli and Posada have attributed some dataset bias to the interpretive influence of managers owing to power imbalances in outsourced data work.[130]

On the persistence of bias, others have noted the presence of discriminatory patterns and representations even in instances where a model was trained on unlabeled or 'cleaned' data. In parsing large volumes of training data, generative models 'discriminate' signal from noise, thus

'learning' dominant patterns as semantic proximities observed in the form of common word combinations or image-text pairs.[131] Following this, Amoore et al. have presented an epistemological view of bias going beyond the "pragmatic terms of direct data-label relationships" to understand bias as immanent to the underlying distribution of the trained model.[132] In this reading, because biased outputs are "not determined in advance by rule, label, or axiom" but instead emerge as a "condensation of probabilities," bias is encoded as a non-representational property of the distribution *independent* of any linear-causal/input-output relation in the data.[133] Wendy Chun similarly describes the encoding of bias through what she calls "proxies": indirect measures or correlations that parallel/index biases in the absence of direct signification.[134]

Amoore et al. thus problematize claims that bias can be rooted out through curation, as discriminatory outputs "may never be traceable since the interstitial samples do not strictly exist anywhere as data points in the training data: they are latent and learned."[135] This capacity of generative models to "dissolve and reconstitute" patterns of hegemony has been vividly illustrated by Offert and Phan in their paper "A Sign That Spells: DALL·E 2, Invisual Images and The Racial Politics of Feature Space." Referencing an experiment revealing OpenAI's vaunted 'de-biasing' of DALL-E2 to consist of simply adding "randomized, hidden keywords at the end of prompts" to circumvent the model's default toward bias, the authors attest to the ultimate "durability" of whiteness as a feature so ingrained in culture as to survive its semantic compression.[136] Yarden Katz, quoting the Black studies scholar George Lipsitz, has also pointed to 'whiteness' as a structuring ideology of data and AI *even when* they do not directly index race: "As the unmarked category against which difference is constructed, whiteness never has to speak its name, never has to acknowledge its role as an organizing principle in social and cultural

relations."[137] It follows then that technical mitigation efforts have been critiqued as surface-level tweaks; doing little to unsettle the politics of discrimination at the heart of data collection and probabilistic knowledge production,[138] and equally imbricated with the politics of defining and generating a *better world*.[139] Additionally, as mitigations largely rely on the voluntary behaviour of profit-seeking companies,[140] observers have also suggested that 'AI responsibility' initiatives serve, first and foremost, to isolate developers from criticism[141] and maintain monopolistic power.[142]

Here again, I contend that Mirzoeff's formulation of visuality presents a unified theoretical framework appropriate to political economic analyses of generative AI. As models do not (strictly speaking) invent their own classes, they necessarily encode the discursive legacies of plantation, imperial, and now counterinsurgent visualization. As such, the models' discriminatory operations enact a "distribution of the sensible"[143] returning bias as the index of what Mirzoeff calls 'Visuality 1,' that is, the deeply-embedded complex of discursive conventions constituting the "coherent and intelligible picture of modernity that allowed for centralized and/or autocratic leadership."[144] In this sense, bias—as the perspective of Visuality 1—overlaps with what Pasquinelli calls the 'eye of the master,' a 'looking' that both enacts and naturalizes capitalist divisions of knowledge, perception, and labour.[145] By distinguishing Visuality 1 from the mechanics of its visualization, an analysis of generative AI through visuality understands bias as a discursive and representational inheritance of the particular *regime of legitimacy* visualized and sustained by the counterinsurgent operation of probabilistic knowledge production, and thus more politically *contingent* than a quality *inherent* to generativity as such. For this reason, I contend that a study of generative AI's political economy in the context of

disinformation and contested legitimacy must attend to counterinsurgency as a mode of visualization and control over and above the content of any specific generation.

The Right to Look: Political Economies of Visualization

The political economy of generative AI is deeply intertwined with those of the infrastructures for collecting, storing, and processing the vast amounts of data necessary for visualization. Differing from the study of AI bias, which centers on the contents and classificatory regimes which make up and structure training corpora, critical studies of the politics of data's production, ownership, and access foreground issues of extractivism, privatization, and governance. As AI's appetite for data grows, social platforms, content hosting sites, and other digital repositories come to serve as primary sites of data collection, often without compensation or the informed consent of the users who generate it. This has given rise to a body of literature critically examining these data practices through the lenses of capitalist extraction, the division and expropriation of labour, and the fundamental asymmetry of knowledge and power which paradigms of proprietary AI both rely on and exacerbate. Following a brief overview of these positions, I conclude this subsection by referring back to visuality so as to read the data politics of generative AI as rooted in the denial of what Mirzoeff calls the *right to look*.

Toward the end of the aughts there began to be exponential growth in the size and capability of machine learning models, despite their theoretical foundations remaining largely unchanged. This progress has been attributed to the availability of large high-quality datasets

along with new methods for splitting up and parallelizing processing tasks.[146] In 2012, researchers at the University of Toronto developed a novel approach to deep learning pioneering the use of GPUs to train classificatory models of unprecedented breadth (in terms of training data and number of parameters) through parallelization.[147] Dubbed AlexNet, the paper is often attributed with pushing the "bigger is better" ethos which has influenced much AI development since, having confirmed the long-standing assumption that model capability could scale logarithmically with increased access to data and compute: "all of our experiments suggest that our results can be improved simply by waiting for faster GPUs and bigger datasets to become available."[148]

While the era of so-called 'big data' has proven a boon for machine learning, danah boyd and Kate Crawford were quick to question the mythology of data's generalization, objectivity, and epistemological value.[149] However, barring some recent detractions[150] and early signals[151] of a shift toward lower resource models,[152] the 'bigger is better' ethos has largely born out to be true, as access to large datasets has measurably improved model performance across tasks,[153] particularly in the case of LLMs, which encode 'general knowledge' from text data scraped from the internet. As an example, the data used to train OpenAI's GPT-4 is rumoured* to comprise over 13 *trillion* discrete data points; resulting in a more than ten-fold increase in model size over its predecessor GPT-3.5.[154] The high rates of improvement observed over the last two decades have relied on the accessibility of large quantities of *high-quality* usable data. Quality and usability have generally been defined according to the "3 V's" of big data: Volume (self-explanatory); Variety (multimodal; diverse; labeled/unlabeled); and Velocity (speed of generation and

* OpenAI does not publish these details.

analysis).[155] Over the same period, social media and content-sharing platforms have emerged as primary vectors for the extraction of high quality user-generated data satisfying each of the aforementioned properties.

With the collapse of the dot-com bubble, the dominant economic model of the internet shifted from one based on selling access (to content, services, and goods) to one based in extraction (collecting and selling user data for targeted advertising, predictive analytics, and now AI training). Referred to as surveillance capitalism,[156] platform capitalism,[157] data capitalism,[158] or vectorialism[159] (among others), accounts of this new business model invariably foreground its figuring of new class relations based on the extraction and control of information from and about users in order to "build predictive models which further subordinate all activity to the same information political economy."[160] Having been discursively transformed from personal property into inert resource by the oft-repeated dictum 'data is the new oil,'[161] political economies of data collection thus become central insofar as they constitute both the principle means of modern capitalist production[162] and a condition of contemporary AI's existence.

Under the extractive model, it is said that platform services come to function as primary collection sites, transforming "smaller and smaller moments of human life"[163] into capital, intensifying the "asymmetric redistribution of power that is weighted toward the actors who have access and the capability to make sense of information,"[164] and supplying the massive quantities of high quality data necessary to train leading generative models.[165] This model has been critiqued from a number of perspectives, with some of the most visible highlighting unsettled legal questions concerning privacy, copyright and licensing, and the (often lack of) informed consent of data subjects.[166] Additionally, the collection of public and private data has also been criticized as presenting a threat to the "foundational and essential openness of the internet," as

increased competition for information and indiscriminate web-scraping incentivizes platforms, websites, and individual users to sequester their data behind passwords and paywalls.[167] Following this, it has been claimed that this has lead to a 'paradox of the commons,' whereby proprietary data and generative models simultaneously close, capture, and degrade the collective informational resources on which it relies.[168] Matteo Pasquinelli has thus characterized the mass extraction of data as an extension of a broader trend in which the capitalist management of labor has "turned all of society into a 'digital factory,'"[169] wherein the shared intelligence of the digital commons is privatized, processed, and sold back as a service.

Through Pasquinelli, we arrive at analyses which understand political economies of data as being rooted in the expropriation and alienation of human labour. In *The Eye of the Master* (2023), Pasquinelli posits a "labour theory of machine intelligence," arguing that AI (and computation, more generally) imitate and emerge from pre-existing capitalist divisions of labour, commodifying 'intelligence' through the abstraction, alienation, and invisibilization of the human labour and relations which produce data and make computation possible.[170] While this dynamic is, of course, at play in the relationship between everyday internet users and the companies training proprietary AI on the content and behavioural data they produce,[171] additional studies have also highlighted the AI industry's reliance on vast networks of so-called "ghost workers"; underpaid labourers tasked with the work of sorting, labelling, and annotating data to be used in supervised training.[172] In each case, sustaining the illusion of AI as the 'automation' of intelligence necessitates the disappearance and alienation of an underlying mass of social labour.

Under these conditions, data 'workers' (both literal and unwitting) have become alienated in the fullest sense of the term, separated from the product of their labour (knowledge; meaning;

visuality) and the means of its production (visualization).[173] Speculating in 1947 on the Automatic Computing Engine, Alan Turing described a hierarchical division of labour separating 'masters,' who control and direct the machine, from 'servants,' who "assemble data that it requires" but have no say in its operation.[174] Describing a decline in what Schumacher called 'free being,' Berry and Stockman similarly address this alienation in Marxist terms, suggesting that these trends further the "proletarianization of the creative production of meaning."[175]

Thought in terms of visuality, the master/servant relation so commonly evoked in technical discourse reflects a dynamic whereby one party claims the ability and authority to visualize reality from the raw data produced by another, who is in turn cut off from the means of visualization—the right to look. Neda Atanasoski and Vora Kalindi have echoed this sentiment, claiming that the exploitation and dispossession undergirding what they call *technoliberalism* reinforces the differential relations enabling the "formation and consolidation of the liberal subject—a subject whose freedom is possible only through the racial unfreedom of the surrogate"—even as it automates servile labour.[176] I suggest there is a clear equivalence to be drawn between the construction and maintenance of this (techno)liberal subject as the "agent of historical progress,"[177] and Carlyle's original conjuration of the Hero as an embodiment of Visuality's authority,[178] whose ability to visualize history similarly relies on a denial of the surrogate's right to look, and thus structuring the politics of visualization and generativity alike. According to Mirzoeff, such visuality presupposes a division of perceptual labour wherein the right to look—to interpret and represent reality—is reserved for authority, while those whose labour supports that work are denied the reciprocal ability to see or be seen as knowing subjects: "for us, there is nothing to be seen." [179]

With this in mind, some efforts have been made to "look back," so to speak. Many academic and artistic projects have sought to critically interrogate the contents, practices, and politics behind the datasets used to train generative models;[180] produce 'counter-data' (actively collecting data representing occluded subject positions, relations, or phenomena);[181] and advocate for or create tools for "opting out" of or otherwise disrupting the collection and processing of data.[182] Another proposal, data cooperatives, promises to rebalance economic and social power, return transparency, and transfer rights of ownership and governance to user/citizens.[183]

While they are steps in the right direction, I suspect that such efforts *misrecognize* the locus of power. Rather, an analysis of generative AI and data through visuality—and specifically the right to look—emphasizes the political and epistemological power of generativity as not strictly (or even primarily) hitched to data in and of itself, but to the mode of its visualization. As attested to by the stubborn and well-documented occlusions and biases of AI, it is not enough to simply be represented in and by data,* and while it is often framed as empowerment, to claim the right of erasure may become akin to exile in a world where AI increasingly dictates legibility. Similarly, while cooperatives offer an alternative structure for data transactions, I struggle to see how this amounts to more than a renegotiation of the terms by which the right to look is sold. If, as Mirzoeff suggests, the power of the right to look is in the claim to *countervisualization*,[184] then a critique of generative AI must, following Mackenzie, specify the abstract nature of its operational power without attributing to it "a mathematical or algorithmic ideality."[185] Projects

* This point has received *some* attention in technical circles, where a few researchers have stressed the need to address the role of model design in conditioning biased or harmful outputs: Sara Hooker, "Moving beyond 'Algorithmic Bias Is a Data Problem,'" *Patterns* 2, no. 4 (April 2021).

to construct “safe” generative models in turn reveal the operational power of generative models as malleable to specific sociopolitical interests, creating methods for circumscribing regimes of legitimacy which diverge from underlying distributions in politically expedient ways.

Counterinsurgency: AI as an Instrument of Control

Studies of artificial intelligence as an instrument of control are situated in the broader historical context of the shift from disciplinary society, as described by Foucault,[186] to contemporary models of power based on the concept of Control. First described by Burroughs[187] and later formalized by Deleuze, the shift to control saw the locus of power move from the enclosures of the institution toward fluid, decentralized networks of continuous surveillance and universal modulation, and thus further integrated into the flows everyday life.[188] This change was also tracked by the sociologist Ulrich Beck who, writing around the same time as Deleuze, theorized the emergence of what he termed the ‘risk society,’ a paradigm of *perpetual uncertainty* in which the neoliberal mode of governance operates according to precautionary logic, increasingly preoccupied with anticipating and managing ‘threats,’ now framed as both inevitable and without definitive resolution.[189] In this context, AI stands to function as a powerful tool for risk management, operating within a *c*ybernetic framework to calculate, manage, and autonomously pre-empt or otherwise counter ‘threats’ with previously impossible scale and efficiency. As such, bodies of critical scholarship have emerged reflecting on the place and effect of AI within a growing ‘surveillant assemblage’ of interconnected technologies,

sensors, and databases;[190] AI's capacity for persuasion and modulatory control; and the rise of so-called "AI authoritarianism."

Advances in AI have allowed security actors to address a critical limitation of the surveillant assemblage first described in 2000 by Haggerty and Ericson, namely the limited processing capacity of the "scattered centres of calculation" where "strategies of governance, commerce and control" are derived from the information collected.[191] As the burgeoning assemblage gathers increasingly more data about subjects, Adrian Mackenzie has observed that accounts of AI's operational power have largely emphasized its promise to "regain control of processes of communication […] whose unruly or transient multiplicity otherwise evades or overwhelms."[192] Because AI-driven surveillance does not only collect and store information, but autonomously generates and acts upon *interpretations* (what Mark Andrejevic has called "*operational surveillance*"),[193] it is thus capable of processing vast amounts of unstructured data and, in turn, shaping the field of visibility/possibility in ways that *pre-empt* threat and reinforce control.

Louise Amoore has also gestured to this capacity, writing that algorithms function as automated mechanisms for "directing and disciplining attention" in security contexts, isolating risks while rendering other patterns peripheral or invisible, and thus "appearing to make it possible to translate *probable* associations between people or objects into *actionable* security decisions."[194] By automating the extraction of patterns and anomalies from otherwise unmanageable volumes of surveillance data, AI effectively transforms surplus information from a governance challenge into a *resource* for pre-emptive and/or modulatory control. Amoore elaborates: "It is precisely this 'connecting of dots' that is the work of the algorithm. By connecting the dots of probabilistic associations, the algorithm becomes a means of foreseeing or

anticipating a course of events yet to take place."[195] In this way, AI models extend what Antoinette Rouvroy et al. have termed *algorithmic governmentality*: a form of autonomous control wherein 'risky' or otherwise 'undesirable' patterns of behaviour are flagged for attention, suppression or redirection.[196] And so it follows that the fantasy of perfect pre-emption should be a central feature of operational surveillance, insofar as it promises the eradication (or rather, successful management) of risk through the combined operations of prediction and just-in-time intervention. However, as demonstrated by the work of Benjamin, Eubanks, and Broussard (among others), these supposedly post-political systems invariably render particular subjects—especially poor and racialized populations—more visible, disproportionately directing scrutiny and intervention onto already marginalized communities.[197]

As Amoore and others have since identified, the origins of algorithmic techniques for visualizing and managing risk lie (or at the very least overlap) with the commercial desire to visualize and pre-empt the consumer.[198] In *The Age of Surveillance Capitalism*, Shoshana Zuboff describes the resulting apparatus as a "sensate, computational, connected puppet" productive of an instrumentarian power for "carrying individuals on a fast-moving current to the fulfillment of others' ends."[199] Building on the work of behavioural economics, the AI-driven apparatus Zuboff calls 'Big Other' influences user/consumer behavior not through direct coercion but the modulation of information environments through automated personalization,[200] content moderation,[201] 'nudging,'[202] and a host of other techniques for structuring behaviour in predictable ways via the deployment of AI as a "persuasive technology."[203] Given mutual affinities of the market and the state, the instrumentarian and pre-emptive powers produced by AI (originally a product of the private sector) thus emerge as the solution to uncertainty preferred by private and state power alike, now locked in alliance.[204] Following this, the literature has also

addressed predictive policing,[205] counter-terrorism,[206] and workplace surveillance,[207] among other applications of AI for anticipatory governance.

In recent years, many "mainstream" accounts of AI as an instrument of control have focused on its deployment in authoritarian contexts. China, in particular, is frequently positioned as the archetype of a rising "AI Authoritarianism" for its development and exportation[208] of AI-based facial recognition, automated censorship, and other data-driven governance schemes, including the often misrepresented 'social credit system.'[209] While these applications are often framed as unique to autocratic regimes, the fundamentally suppressive logics of algorithmic governmentality are also observed in liberal-democratic states, where it is more often mediated through the policies of private companies, particularly social media platforms,[210] whose economically-driven techniques of surveillance and modulation function as a "valuable proxy for authoritarian control."[211] In light of this, narratives about AI authoritarianism and the "China threat" have been presented as largely functioning to support securitization and the development of AI that is "safe" or otherwise compatible with liberal-democratic governance.[212] As such, so-called AI authoritarianism and liberal-democratic techniques of algorithmic governance thus operate through parallel mechanisms, though in the West this control is typically justified via its discursive distancing from state authority and appeals to risk management, harm reduction, and the preservation of democratic order against existential threats.

In *Do Artifacts Have Politics?* Langdon Winner paraphrases Engels, asserting that "the roots of unavoidable authoritarianism are […] deeply implanted in the human involvement with science and technology."[213] That is, Engels saw the centralization of authority as a *requirement* of increasingly complex technical systems, the adoption of which in turn requires the "creation and maintenance of [authoritarian] social conditions as the operating environment of that

system."[214] In this light, regimes of *operational surveillance* or the rise of "AI authoritarianism" can not be adequately explained as the product of coincidence, misadventure, or malice, but rather as the inevitable effect of AI's integration into processes and infrastructures of communication, knowledge production, and governance, even regardless of the specific political economic contexts of this deployment.

As discussed earlier in this review, Winner differentiates two ways in which a technology can be understood to contain political properties: as an *inherent* quality, wherein its technical and political operations are synonymous; and through its *intentional* application as a means of "settling an issue in a particular community."[215] As a strategy of anticipatory governance, where the goal is not merely to react to threats but to pre-emptively neutralize them and stabilize a regime of legitimacy through environmental and behavioural modulation, I believe that counterinsurgency presents a useful model for apprehending the internal and intentional dynamics shaping AI as both a technical apparatus and political project of control. Generative models operate as cybernetic systems, refining their outputs through feedback loops integrating user behaviour, adversarial intervention, and moderation/alignment constraints. This recursive process also reflects broader political-economic dynamics, as AI governance is not merely rule-based but continually shaped by platform incentives, regulatory pressures, and the shifting politics around the definition and policing of legitimate information and behavior. This provides a useful lens for understanding the "safe" generative model as both a product and self-regulating mechanism of counterinsurgent visuality. Thus, I contend that AI models exhibit and extend counterinsurgency as an inherent quality (as they autonomously visualize and enforce a regime of legitimate patterns) and through that quality's intentional deployment in specific security

contexts, where the model's visualization directs attention and real-world applications of force aimed at identifying and pre-empting anomaly—however defined.

As an intensification of counterinsurgent visuality, deployments of AI for control epitomize the risk society's prioritization of technocratic population management over social and political amelioration. Perhaps better described as *stability operations*, such techniques of algorithmic governance function not to resolve underlying crises but to manage their effects, modulating the social in service of equilibrium. Regarding its operationalization in the context of 'disinformation' and the crisis of legitimacy, the instrumentarian power of "safe" AI to identify and *disaggregate* statistically-significant patterns of insurgency thus appears particularly suited to the task of reauthorizing Visuality 1 in a heterogeneous field of information and politics increasingly characterized by confusion, uncertainty, and distrust.

Critical Disinformation Studies

While 'disinformation'* has long been a feature of political communication, the twin political upsets of 2016—the Brexit referendum in the UK and Donald Trump's victory in the US Presidential election—catalyzed a surge of academic and policy debate about the role of social platforms and other networked media in shaping public discourse. In the years following, artificial intelligence—particularly generative AI—has emerged as a focal point for intensifying

* While definitions of disinformation, misinformation, malinformation, and "fake news" overlap and vary across academic and policy circles, for the purposes of this dissertation, disinformation is adopted as an umbrella term, denoting any information considered harmful or disruptive to social order and therefore in need of remedial intervention.

concerns about ‘epistemic integrity’ and the political influence of misleading or deceptive content. However, scholarship on disinformation, the role of AI, and proposed countermeasures is roughly split. While ‘mainstream’ Disinformation Studies (here defined as research rooted in the assumption that current disorder is attributable to structural transformations of the public sphere, particularly social media and its role in “driving political polarization and the prevalence of disinformation”)[216] generally frame AI as both cause and exacerbation of declining trust and epistemic decay, Critical Disinformation Studies question the core assumptions of mainstream narratives, analyzing the discourses and political economic factors shaping how disinformation is defined, studied, and responded to.[217]

Disinformation Studies & AI

Post-2016, mainstream narratives of disinformation typically emphasize the effects of the internet and social media, linking access to information, peer-to-peer communication, algorithmic recommendation, and ‘bots’ to waning institutional trust and the growing spread and political salience of false or misleading information.[218] Under this framework, artificial intelligence is presented as a ‘force multiplier’ such that, if left uncontrolled, disinformation operations which either exploit or employ AI pose existential threats to liberal-democratic society.[219] In turn, these narratives are marked by a predominance of martial rhetoric which does everything but understate its stakes, as when Renee DiResta (a prominent advocate of the disinformation narrative) described the issue as “a Hobbesian information war of all against all” leading to a “tactical arms race” with platforms and AI at its centre.[220]

On the whole, the literature describes two primary ways that social media platforms and the algorithms which govern them contribute to the problem of disinformation. For one, platforms and AI-driven recommender systems are described as spreading mis/disinformation as a matter of course, owing to the economic incentives of surveillance capitalism and a confluence of technical and social factors privileging the spread of emotionally-driven (but ultimately false) content.[221] Second, algorithmic techniques of micro-targeting and behaviour modification are said to be intentionally exploited for the purpose of coordinated influence operations aimed at effecting social and political outcomes in largely liberal-democratic societies.[222] Common to the field, this theory of causation frames the influence and spread of "problematic" content as a downstream effect of structural changes in the information infrastructure and the actions of "malicious stakeholders" to unsettle "consensus reality" through the creation and/or exploitation of "filter bubbles."[223] Despite a growing body of contradictory evidence (discussed later), adherents to mainstream narratives of disinformation persistently lay the blame for "democratic backsliding" squarely at the feet of social media platforms, who researchers and policymakers continue to criticize as ineffective and unwilling partners in discourse management.[224]

Disinformation studies present generative AI, in particular, as a death knell for the integrity of digital communications. In their seminal 2018 paper "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," Chesney and Citron outline a panoply of individual, institutional, and systemic threats that generative AI pose to liberal societies, emphasizing the technology's capacity to deceive, distort democratic discourse, manipulate elections, exploit division, and erode institutional trust.[225] While much of the literature foregrounds the ability of generative AI to create convincing 'fakes,' this concern often extends beyond discrete instances of deception to the broader epistemic consequences of

synthetic media.[226] To this point, researchers emphasize how the mere possibility that content may be AI-generated "contributes toward generalized indeterminancy and cynicism,"[227] cultivating skepticism and dismantling what Regina Rini, describing recordings' role as "acute correctors of the testimonial record," has called the *epistemic backstop.*[228] Megan Hyska has offered a complementary perspective, arguing that generative media compromise the capacity to communicate through acts of *showing*—practices that, she suggests, uphold "relational equality through asymmetries in communication, and [as a result] express a distinctive sort of respect for the audience."[229] Insofar as generative AI simulate the form of testimonial showing while severing it from its grounding, Hyska suggests they contribute to the erosion of testimonial trust from another angle. As a secondary effect, this situation is also thought to increase the perceived credibility of textual disinformation, now evaluated within "an information ecology where different competing truth claims are presented in a cue-rich and seemingly authentic format alongside accusations of disinformation."[230] To this end, numerous studies document LLMs' ability to 'supercharge' adversarial information operations,[231] autonomously generating and distributing high-volumes of targeted disinformation in the form of news articles and social media interactions to simultaneously promote and create the illusion of consensus around strategic narratives—a tactic called *astroturfing*.[232]

Citing the unsettling of evidentiary value and ability to "flood the zone" with false or misleading information,[233] much of the commentary on disinformation and generative AI prognosticates an immanent *infocalypse*, a "fucked up dystopia"[234] where the semantic entropy of proliferating fakes brings about the "end of reality as we know it."[235] Perhaps unsurprisingly, then, disinformation became (and, in some sectors, remains) a well-funded area of research following the political upsets of 2016 and emergence of highly-capable generative tools,

producing a flood of reports, analysis, and countermeasures aimed at understanding and curtailing online disinformation. This work has in turn given rise to an array of proposed solutions to the problem of disinformation.

Generally corresponding to a handful categories identified by Farkas and Shou,[236] typical responses include: ‘Policing the truth,’ referring to the identification, punishment, and removal of disinformation by administrative or juridical means (i.e. the revision and enforcement of platform content, usage, and labelling policies;[237] proposed changes to Section 230 of the US Communications Act,[238] etc.); ‘Technological solutionism,’ which assumes that new IT tools—such as automated content moderation,[239] algorithmic de-amplification,[240] and detection[241]—will inherently contribute to solutions (this perspective also extends to the development of "safe" generative models); and ‘Re-establishing centers of truth-making," describing efforts aimed at “reclaiming the authority and capacity of certain fields to once again produce and speak on behalf of Truth,”[242] building societal ‘resilience’ against disinformation by employing third-party fact-checkers,[243] encouraging ‘prompted reflection,’[244] implementing ‘pre-bunking’ or ‘inoculation’ strategies,[245] and adopting cryptographic provenance schemes like the C2PA.[246] Critically, these approaches reflect a broader tension underlying the field of disinformation research and governance, where AI is understood as fundamentally dual-use[247]—both a persistent threat to information integrity and the keystone of an emerging defensive framework designed to uphold it.

Ultimately, mainstream disinformation studies tend to construct the problem as one of technological misadventure and malign influence, prioritizing solutions that rely on AI-based moderation, platform governance, and regulatory interventions to ‘restore order’ to the digital public sphere. This technocratic approach both reflects and promotes an effectively *depoliticized*

understanding of disinformation, treating it as a product of technical failure rather than a symptom of deeper societal fractures. In centering technical fixes, this approach often overlooks the underlying social, political, and economic conditions that make disinformation compelling in the first place—a critique to which critical disinformation studies directly respond.

Critical Perspectives on Disinformation and AI

Although not explicitly named before 2021, Critical Disinformation Studies has emerged as a distinct subfield of disinformation research. Critical studies challenge the prevailing assumptions underlying mainstream narratives, particularly the emphasis on disinformation as a novel technical failure or the product of interference on the part of variously defined 'bad actors.' Rather than treating it as a form of informational warfare or pollution—something to be identified, removed, and mitigated through technical interventions—critical scholars approach disinformation and the discourses surrounding it as modes of knowledge production reflecting broader struggles over power, legitimacy, and order in the public sphere. As Kuo and Marwick argue in their pivotal article "Critical Disinformation Studies: History, Power, and Politics," understanding disinformation requires moving beyond normative frameworks to examine how knowledge is being contested, circulated, and weaponized, particularly from BIPOC, queer, feminist or otherwise marginalized social and political positions[248]—an exercise overlapping the study of visuality which runs through this dissertation. As such, this perspective shifts the focus from identifying technical causes and solutions to considering the social and political economic

conditions which shape disinformation, while also interrogating how disinformation discourses themselves function to regulate dissent and reinforce dominant regimes of legitimacy.

A core tenet of much critical disinformation scholarship is a disagreement with popular attributions to recent shifts in technological or political conditions. In *The Disinformation Age*, Bennett and Livingston provide a representative account of this position, linking the current state of information 'disorder' to a breakdown of institutional authority wherein traditional gatekeepers like professional media, political parties, and peer-reviewed science have become increasingly ineffective anchors of public debate.[249] In explanation, the authors present a historical account of declining public confidence, framing it as the result of "decades of capture and erosion of governing institutions by wealthy interests and aligned political elites, unable to sell their actual agendas to the public without increasing levels of disinformation."[250]

This history begins in the early-to-mid 20th century with the re-branding and mobilization of 'propaganda' for democratic management, now called 'public relations.' Described by Edward Bernays as the "executive arm of invisible government,"[251] public relations names the informational strategies by which Western elites sought to prevent political and economic disruption through the "scientific"[252] management of public opinion. Bennett and Livingston's account characterizes the period between World War II and the 1980's as relatively stable, where the established framework of "managed communication" effectively counterbalanced the episodic manipulation with the pretense of carrying authoritative information, resulting in high degrees of trust in the institutions of politics and the press.[253] However, the growing influence of neoliberal economics over the second half of the century saw a shift toward more *systemic* deception, as think tanks, private foundations, and aligned politicians and media companies produced increasing volumes of 'official disinformation,'

manufacturing legitimacy for policies that would otherwise struggle to gain democratic support.[254] As Kuo and Marwick have noted, these strategies more often than not exploit the "deep frames" of racial and economic antagonism to justify the hollowing out of public services, as with the Reagan campaign's regular invocation of the Africa American "welfare queen" taking advantage of hardworking (white) Americans,[255] and any number of examples since.

Bennett and Livingston note that, following initial success, neoliberal policies eventually became harder to sell to voters who bore the brunt of their negative socioeconomic outcomes, claiming that "such results reflect the basic contradiction in the neoliberal project: people would not buy it on its own terms."[256] In reaction, aligned corporations, politicians, and institutions deployed a spiral of increasingly implausible and/or unprovable justifications alongside a host of political strategies designed to insulate neoliberal policy agendas from democratic intervention.[257] The critical narrative put forth by Bennett and Livingston proposes that the last fifty years' blatant undermining of democratic institutions by ideologically-captured politicians, experts, and media has directly led to the crisis of public trust in official information at the centre of the current disinformation order. This erosion of trust, they argue, primarily stems from prolonged exposure to the strategic narratives and performative rhetoric issued by once-credible authorities. As this messaging proliferated—from politicians, spokespeople, and complicit media—it bred a deepening skepticism toward both the content and the institutional channels of so-called 'official' communications.[259]

This perspective on the origins of the crisis directly contradicts that of mainstream disinformation studies, which tend to instead place the blame for epistemic disorder, political polarization, and the electoral successes of right-wing political parties squarely on the shoulders of irresponsible digital platforms and the manipulations of bad actors. As Bennett and Livingston

argue, the crisis reflects a breakdown of institutional legitimacy with much deeper roots than is accounted for in mainstream discourse, a contention supported in part by polls indicating downward trends in institutional trust over a timespan pre-dating the emergence of the internet and social media.[260] Peter Dahlgren, writing in 2018, has similarly linked growing distrust in the institutions of politics and media to historical trends whereby "deception by power elites had become more sophisticated and systematic for many decades, and citizens of all persuasions had become increasingly aware that politicians, lobbyists, corporations and other actors engage in spin, disinformation and even lies."[261] According to this understanding, the rise of right-wing populism is also explained as an unintended consequence of the neoliberal practice of selling policy through division and the hollowing out of democratic institutions, creating a political vacuum wherein the concerns of citizens are increasingly without mainstream political representation—a situation readily exploited by political reactionary's claiming to represent ordinary people.[262] This narrative thus suggests that efforts to combat disinformation without addressing the deeper roots of the legitimacy crisis risk entrenching the very dynamics that fueled its rise.[263]

As the above perspective illustrates, a defining characteristic of critical approaches is the rejection or critique of core assumptions underlying mainstream narratives of disinformation; chief among them being the fallacy of an 'age of objectivity' which preceded the current state of disorder. In a 2021 article for *Harper's*, Joseph Bernstein condenses this attitude, opening with a line comparing the 20th century paradigm of mass media to Creation, an order ordained by God, and hence right:

> In the beginning, there were ABC, NBC, and CBS, and they were good. Midcentury American man could come home after eight hours of work and turn on his television and

> know where he stood in relation to his wife, and his children, and his neighbors, and his town, and his country, and his world. And that was good. Or he could open the local paper in the morning in the ritual fashion, taking his civic communion with his coffee, and know that identical scenes were unfolding in households across the country.[264]

With a cheeky reference to Genesis, Bernstein underscores the nostalgic longing that is so characteristic of mainstream disinformation discourses, which frame the present as a catastrophic departure—or fall from grace—from the "consensus reality" of a more stable and truthful past.[265] Plagued by confirmation bias and wielding the bullhorn of social media, it goes, newly empowered publics "are busy shattering the edifice of liberal-consensus reality into a million little pieces, with no hope of any universal project on the horizon that might be capable of reassembling its fragments."[266] Rather than recognizing the relative coherence of the 20th century mediascape as an ideological construct maintained through public relations, institutional control, and political gatekeeping[267]—"the product of social institutions that were once so powerful they could hold together a shared picture of the world"[268]—mainstream discourses frame its unsettling as a crisis of knowledge itself—an affront to the order of things.[269]

Critical perspectives, on the other hand, suggest it is not so much the loss of *truth* that is mourned, but the decline of an order in which political and economic elites had long monopolized the levers of democratic management. In this way, critical literature presents contemporary disorder as the latest of many historical transformations in public and political communication. In *The Structural Transformation of the Public Sphere*, Habermas traces a series of changes in the organization and role of the public sphere, which he conceived as a space of rational-critical debate wherein 'the public' (the people, aestheticized as a political subject) could engage authority "in a debate over the general rules governing relations."[270] Habermas' account

tells of a shift from the bourgeois ideal of deliberative democracy and public checks on power* to the mass-mediatization of the public sphere as a forum wherein authority is increasingly insulated from public scrutiny.

More recently, Jungherr and Schroeder have adapted Habermas' thesis to suggest that current concerns over disinformation extend from structural transformations of the public sphere brought about by the internet. Where once access to the public sphere was managed by a handful of gatekeeping institutions, this ability has been significantly disrupted by the migration of public debate and information sharing onto digital platforms, and the proliferation of actors and increased visibility enabled by platforms has diluted established controls over agenda-setting and the regulation of acceptable political discourse, resulting in a destabilizing "shock" to those previously dominant arbiters of authoritative information.[271] Though Jungherr and Schroeder acknowledge this transformation as fundamentally destabilizing (using the term "public arena" to better describe it as a site of contestation), they suggest this is an effect of technology's capacity to reveal political and epistemological perspectives otherwise unaccounted for in the public sphere (citing a lack of representation by established parties, movements or interest groups†) and providing opportunities for previously disaggregated subjects to "find each other, distribute information, coordinate, and organize their challenge to the status quo."[272] Extending this notion, Bauer and Nadler have suggested that the problematic of disinformation itself

---

* Like the fallacy of prior objectivity, Habermas' description of the bourgeois public sphere as open and inclusive has since been critiqued as a fundamentally hegemonic space marked by the exclusion of women, the poor, and racialized citizens. See: Nancy Fraser, "Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy," *Social Text*, no. 25/26 (1990): 56, https://doi.org/10.2307/466240.

† In 2022, Habermas also took up the fundamental transformations wrought by digital media, albeit while remaining normatively committed to more universal principles. Jürgen Habermas, "Reflections and Hypotheses on a Further Structural Transformation of the Political Public Sphere," Theory, Culture & Society 39, no. 4 (July 1, 2022): 145–71.

"*reintroduces* democratic potential" by sparking public debate over the design and control of communication technologies and indexing the deep political schisms underlying current crises.[273] In light of this, critical disinformation asks: what interests are served by prevailing disinformation narratives?

Jungherr and Schroeder, among others, have observed that rhetoric about a "threat to democracy" seems to resonate most strongly among (generally centrist) institutions and elites whose influence is challenged by structural transformations of the public arena.[274] In this view, these organizations' conceptualization of the problem of disinformation in effect constructs it as a discursive mechanism for restoring hierarchies of power that once governed the public sphere in their interest.[275] It follows, then, that many of the same actors and institutions that facilitated the crisis have also played a central role in funding and directing disinformation research after 2016. As such, critical scholars have drawn attention to the role of legacy liberal institutions—The Aspen Institute, Ford Institute, New York Times, Council on Foreign Relations, Atlantic Council, and the World Economic Forum, among others—as primary funders (and definers) of disinformation research.[276] It has been noted, however, that this relationship is not fixed, and even as liberal-democracies undergo a shift in the balance of political power, terms like "fake news" and "disinformation" act as *floating signifiers*, invoked in equal measure by rival hegemonic projects, each "seeking to provide an image of how society is and ought to be structured."[277] However, Kuo and Marwick again remind us of the fundamental "whiteness" of hegemony and the role that disinformation narratives play in reproducing and protecting "historically rooted systems of power that privilege white perspectives,"[278] a realization with broad implications for the study of disinformation in the Global South, where Western-centric frameworks often fail to account for local dynamics.[279]

In light of this, critical scholarship often frames contemporary anxieties around disinformation—and, more recently, generative AI—as a *moral panic*.[280] As formalized by Stanley Cohen[281] and Stuart Hall,[282] moral panics emerge from the definition of a given issue as a "threat to societal values or interests"—positioned as the *folk devil*, a "visual reminder of what should not be."[283] This arousal of public concern by panic actors (often established interests) thus functions to authorize new laws or other countermeasures aimed at restoring order through the folk devil's defeat or neutralization. A defining characteristic of moral panics is the exaggeration of the actual threat posed by the folk devil, exemplified in the positioning of digital platforms and generative AI as objects "onto which a broad array of hopes and fears is projected and envisioned as a potential solution to, or possible problem for, the world at large."[284]

In supporting the moral panic claim, critical scholars point to a growing body of empirical evidence indicating that the threats of disinformation and generative AI are overstated.[285] With regards to disinformation, this research suggests that it makes up only a small fraction of what people consume;[286] isn't disproportionately shared;[287] is generally seen as less credible;[288] and its impact on major political events is often exaggerated[289]—throwing up questions as to whether digital disinformation is as novel or threatening as is presented.[290] Here it is important to note the influence of likely discrepancies in how disinformation is defined by "panic actors" and the authors of the above studies, as well suggestions that the persistence of disinformation narratives in the face of contrary evidence betrays an operationalization of the term against anti-establishment narratives and political organizing. To this point, Tim Hayward has critiqued anti-disinformation research by Stanford and NYU (with support from the US Cybersecurity and Infrastructure Security Agency), citing the *Virality Project*'s labelling of dissenting but otherwise credible information as 'disinformation' as evidence of a research trend

wherein political authority often displaces "epistemic authority as the basis for identifying disinformation."[291]

With regards to generative AI, critical scholarship also contends that fears about the novel epistemic effects of the technology are both overblown and misdirected.[292] Simon et al. observe that "the problem is not that people do not have access to high-quality information but instead that they reject high-quality information and favor misinformation," reflecting the view that, while generative AI may increase its quality or *supply*, there is little evidence they meaningfully effect the visibility or level of *demand* for disinformation.[293] Relatedly, critical studies also highlight the fact that concerns about media manipulation and public distrust—such as the so-called 'liar's dividend'[294]—long predate[295] the emergence of generative AI, and are more accurately attributed to partisanship and eroded institutional trust than to any effect of the technology itself.[296] In sum, consider Joshua Habgood-Coote's suggestion that, by focusing on disinformation and generative AI as novel technological problems, apocalyptic narratives about generative AI sideline the opportunity to reassess "the management of our practices for producing and receiving [information]" in favour of technical solutions defending the status quo.[297] In this way, moral panic over the folk devils of disinformation and generative AI primarily function to authorize extraordinary information controls on the pretext of restoring order.

It is important to note that many of the authors discussed here do not deny the reality of threats posed by real existing disinformation. Rather, critical disinformation encourages attention to the possibility of "legitimate concerns over real but marginal phenomena overshadowing the very real benefits of increased access to public arenas" and leading to closure.[298] To this end, the inducement of moral panic around disinformation and generative AI has been read as a

securitizing move,[299] deploying the hyperbolic rhetoric of existential threat in order to create states of exception wherein "extraordinary measures" can (and must) be put in place.[300] In a 2020 report for *Data & Society* on the securitization of 'fake news' by Malaysia's ruling coalition ahead of that country's 2018 general election, Gabrielle Lim also points to the securitizing effects of disinformation discourses in the US, UK, and Canada[301]—a function complimented by apocalyptic rhetoric about the dangers of 'uncontrolled' generative AI.* In light of this, scholars have turned their attention to a critique of the "proactive measures" that this securitization legitimates.

These measures have been criticized for entailing forms of coordination between state and corporate actors that often blur the line between governance and surveillance. Renee DiResta—formerly of the Stanford Internet Observatory, and a key figure in shaping securitizing narratives around disinformation and generative AI—has suggested that confronting the problem of disinformation and 'harmful content' demands a "new and functional defensive framework," envisioned as a tightly integrated security apparatus spanning public and private sectors.[302] In parallel, others have called for a return to the "epistemic paternalism" of authoritative gatekeepers, reinforcing the position of platforms, experts, and state actors as the arbiters of trustworthy information online.[303] Such collaborations are said to reflect and reinforce neoliberal logics of 'technosolutionism,' where sociopolitical problems are reframed as technical challenges solvable through enhanced techniques of surveillance and moderation.[304] These efforts have raised concerns about the erosion of free expression and democratic discourse

* As a note, I believe the securitization of AI to be a much more subtle and easily sold means of controlling discourse than doing the same for disinformation. By securitizing the technology, not the discourse, security actors are able to expand controls under the supposedly apolitical auspices of negating the disastrous effects of "dangerous AI."

online,[305] with the *a priori* exclusion of 'disinformation' from the public arena being likened to censorship via "prior-restraint,"[306] and framed as potentially chilling political speech on mainstream platforms—especially for marginalized or dissenting voices.[307] Meanwhile, others have highlighted a global trend of governments exploiting anti-disinformation policies to suppress or criminalize legitimate speech, extending the logic of securitization beyond platform governance to broader means of political repression.[308] Finally, integrating insights from critical AI, Tarleton Gillespie has also questioned the utility and ethics of granting automated systems increasing epistemic authority, noting how moderation algorithms obscure the value-laden decisions they encode, and exemplify a broader tendency to "paper over" rather than confront incommensurate values.[309]

Gillespie's point about depoliticization resonates with what some see as a central critique of mainstream disinformation studies. With the rhetorical question "When the underlying assumption is that disinformation is a threat to democracy, what is 'democracy'?" Lenoir and Anderson contend that perspectives which frame disinformation as a problem to be *eliminated* share an assumption that democracy is defined by stability and consensus rather than disagreement and contestation.[310] They suggest that, in doing so, such efforts risk erasing conflict by treating it as pathological, rather than constitutive of deliberative democracy. As they put it, the prevailing Western approach to disinformation governance increasingly resembles a project to "make conflicts disappear," or at least deny their necessity to democratic life.[311] With their question, the authors distill a broader concern; namely that attempts to control the flow of information in the name of protecting democracy suppress the very contestation that defines it, further subordinating institutions to elite control and initiating a "drift toward authoritarianism" in democratic societies around the world.[312]

In this reading, authoritarianism rises not only in response to political or economic conditions, but as an inevitable result of elite attempts to "impose coherence onto a fractured world," an impulse shared by both liberal establishments threatened by 'disinformation' and the right-wing reactionary movements which oppose them.[313] The crisis of legitimacy and attendant emergence of 'disinformation' and rising populism as subjects of concern thus index a breakdown of previously effective techniques of democratic management. Rancière helps clarify the stakes, writing that the function of 'populism' in elite discourse is that of conceptual placeholder: the "convenient name" which simultaneously identifies and delegitimizes the manifestation of popular dissensus with authority.[314] In Rancière's formulation, populism marks not a deviation from democracy, but a confrontation with the struggle over the identity, voice, and desires of 'the People,' and the crisis of reason or truth is instead revealed as a crisis of governance—one expressing "the intense wish of the oligarch: to govern without people […] to govern without politics."[315] In this view, the fight against disinformation becomes a project of depoliticization—one that seeks to neutralize conflict rather than engage it, and in doing so, aligns with broader intensifications of post-political governance. This technophilic treatment of symptoms, rather than causes, at best results in a *stalemated peace*, a concept from international relations describing a situation in which violence is minimized, but underlying sociopolitical issues go unresolved.[317] As Oliver Richmond has noted, this model of peace increasingly depends on "new forms of counter-insurgency and stabilization in the context of growing global polarization," transforming dissent from democratic expression, to a threat to be managed[318]—a claim echoed by Bernard Harcourt and Kristian Williams, who both argue that liberal governments increasingly regulate domestic populations via counterinsurgency,[319] as well as a

handful of scholars who have applied counterinsurgency to a reading of disinformation countermeasures.[320]

In 2015, Jayson Harsin described an ongoing shift from the relatively stable circulation of disciplinary truth, to emergent *regimes of posttruth*, marked by a highly-fragmented 'truth market' and corresponding to societies of control.[321] In this framing, disinformation narratives become central to the maintenance of political and epistemic authority via proliferating 'truth games,' tactical performances wherein resource-rich elites attempt to manage the terms of appearance and participation in order to "block the emergence of more inclusive social justice agendas or even the reorganization of the place of political agency itself"[322] —counterinsurgent disaggregation in all but name. Through this we arrive back at a core contention of this dissertation: that the strategy of disaggregation—the pre-emptive dispersal, containment, and deflation of oppositional publics—best captures the political operation of "safe" generative models under the banner of epistemic integrity and democratic resilience.

Taken together, these critiques reveal that what is often framed as a technical crisis of truth is better understood as a political crisis of authority—one in which the fight over 'disinformation' serves to define and enforce the boundaries of legitimacy. Ultimately, Critical Disinformation cautions scholars and commentators against fanning the flames of moral panic and uncritically taking up the mantle of a centrist campaign against insurgency such that the realization of "a more open public arena through digital media becomes synonymous with disinformation and the decline of democracy."[323] I work with this in mind, approaching generative AI as a contested site where competing claims to authority, legitimacy, and world-making play out. In tracing how these technologies are being secured against insurgent

potentiality, we apprehend not only how visuality intensifies, but how it might yet be re-imagined.

Conclusion

On the whole, I situate this dissertation as a unique contribution to the growing body of critical disinformation scholarship, presenting a cross-disciplinary framework for apprehending the "safe" generative model as a technopolitical device developed and deployed in the context of crisis. In this reading, these models function not merely as risk mitigation, but as means of pre-emptive political containment—mediating and organizing information and subjects in ways that align with broader techniques of counterinsurgent governance. With this work, I seek to foreground the political stakes of "safe" generative media, and to expand the analytic vocabulary available to scholars confronting the technocratic terms of contemporary epistemic and political authority.

In framing "safe and responsible" AI as an intensification of counterinsurgent visuality, I also build on and extend current critical AI scholarship. While existing work has effectively unpacked how generative models visualize according to a statistical (non-representational) order of knowledge with discursive roots in plantation and imperial visuality, I feel it has largely stopped short of fully theorizing the models' underlying logic as *counterinsurgent*: an analytic turn necessary to understanding their political-economic utility as instruments of stabilization. As a result, even critical efforts to mitigate bias and other unfair classifications risk reproducing counterinsurgency as an inherent political logic. By addressing the politics of representation

without confronting AI as a fundamentally counterinsurgent apparatus, such reforms risk merely recoding classification under new terms of legitimacy—changing the face of control, not its function. This difficulty stems, in part, from the paradoxical nature of counterinsurgent visualization, insofar as it visualizes and generates its legitimacy via a non-representational (or invisual[324]) logic of governmentality. This dissertation extends the insights of critical AI scholarship by working across political economy, critical AI, and disinformation studies, unpacking a complex of regulatory, technical, and epistemic closure constructing generative AI as a technical and discursive intensification of counterinsurgent visuality. Taken together, closures produce "better" or "safer" models through alignment with the strategic imperative of *disaggregation*: the containment and dispersal of insurgent publics to undo and prevent their coherence as oppositional political subjects. This work thus departs from critical AI not by rejecting its insights, but by extending them—going beyond epistemological critique to theorize how generative AI are enrolled into a broader project of post-political counterinsurgent control.

I feel it is important, here, to acknowledge that academic critiques of mainstream disinformation narratives and countermeasures at times overlap in uncomfortable ways with positions advanced by commentators who push back against what they call the "censorship industrial complex," referring to an alliance of political, economic, and media institutions coordinating the silencing of conservative voices.[325] In turn, the censorship thesis has been criticized as a reductive framework reliant on exaggeration and misrepresentation, veering perilously close to conspiracy, and playing into the hands of autocrats by discrediting genuine anti-censorship concerns.[326] Though it would be easy to fall into the trap of either of these camps, I believe my employment of visuality as a theoretical framework establishes a degree of critical distance from these totalizing frames. While my work confirms the existence and

influence of censorious actors, my analysis ultimately contends that outright censorship (a partisan action) is a secondary effect of broader discursive and structural intensifications of counterinsurgency as the dominant mode of visualizing, naturalizing, and managing legitimacy—an agnostic operation that is more procedural than partisan.

I believe a study of disinformation and generative AI via the crisis of counterinsurgent visuality thus allows us to speak of and render visible its *intensification* in the artifact of the "safe" generative model without recourse to conspiracy or degrees of conjecture otherwise not appropriate to a doctoral dissertation. In this sense, visuality functions not only as a theoretical framework, but also a methodological orientation—as the under-articulated, but operative, thread across political economy, critical AI, and disinformation studies. At times, this dissertation adopts a historically and genealogically informed frame, not to offer its own history or Foucauldian genealogy, but to articulate the conditions under which generative AI emerges as an intensification of visuality and power. Following Bratich—quoting Jason Read—my project applies visuality to a *symptomatic reading* of the "unstated presuppositions and problematics"[327] that structure claims about 'safe' AI and epistemic disorder, and that ultimately produce/authorize the deployment of generative AI as a counterinsurgent apparatus. In using visuality to surface the logics of control and intelligibility embedded across the design, deployment, and narration of generative AI, I take seriously Foucault's proposition that power is better understood as a strategy without a strategist—a distributed rationality embedded in practice, rather than a centralized or conspiratorial force.[328] In this reading, generative AI's 'conscription' into intensified counterinsurgency unfolds without the presence of a singular, scheming subject, while still being nonetheless *intentional*. What emerges is not a conspiracy but

a cascade: a convergence of regulatory, technical, and epistemic conventions guiding generative AI into the service of counterinsurgent control.

This dissertation is unique in its application of visuality as a unifying theoretical framework, tracing a connective thread across political economy, critical AI, and disinformation studies. In doing so, it diagnoses how generative AI operates both epistemologically and operationally within an intensifying regime of counterinsurgent visualization—one shaped by discourses of disinformation, existential risk, and safety. In response to the crisis of legitimacy and the aggregation of insurgency, "safe" generative models are emerging as technical instruments of *disaggregation*: inhibiting organization, enforcing legitimacy, and further denying the right to look. Here, visuality proves especially well-suited as a frame, being, in Mirzoeff's formulation, a fundamentally disaggregative force itself: "[V]isuality separates and segregates those it visualizes to prevent them from cohering as political subjects."[329] The 'safe' generative model, then, can be understood not as an aberration, but as a continuation—emblematic of a centuries-old project of visualization and power, updated and intensified for a new era. This study offers an account or diagnosis of that intensification, the futures it portends, and raises the open question of whether counterinsurgent media might yet be reappropriated to visualize emancipation.

PART ONE
Visuality: Genealogies of Media and Power

Introduction

This first part of the dissertation functions to develop the theoretical and historical foundation for the following parts by situating generative AI within broader genealogies/histories of media and power. Following a framing of how visualization operates as a means of organizing perception and constraining meaning, the section introduces visuality as the dissertation's core analytic frame, adopting Nicholas Mirzoeff's formulation as a mode of 'seeing' that does not merely *r*epresent the social world, but both constructs and governs it through complementary operations of classification, separation, and aestheticization. As a historically contingent formation, visuality "sutures authority to power," delineating not only who is authorized to 'see,' but how and under what conditions subjects are rendered visible.[1] Rather than treating it as a static regime, Mirzoeff offers a genealogy of visuality's adaptations (its *intensification*) in response to challenges from below—specifically the subaltern's claiming of the right to look and *countervisualize* its reality, understood as the performative assertion of the autonomy to self-visualize.

Through this lens, I construct a compound genealogy wherein media technologies emerge not simply as means of communication or documentation, but as material infrastructures of visuality—affording new ways of ordering the social, and in turn authorizing, contesting, or intensifying particular configurations of material and discursive power. In order to illustrate this,

I construct a periodization framework correlating Mirzoeff's historical complexes of visuality (plantation, imperial, and military-industrial or counterinsurgent) to parallel developments in both media (from inscription, to index, to data) and modes of governance/power (sovereign, disciplinary, and control). By foregrounding how each complex of visuality is entangled with specific media forms and political rationalities, I seek to illustrate how media technologies function as both agents and intensifiers of visuality and power—enacting and naturalizing particular ways of visualizing and governing the social, while also affording new opportunities for resisting that authority. That is, media technologies and their affordances are themselves historically configured by, and actively configure, the operations of visuality.

From the mapping and cataloguing practices of early modern imperialism to the emergence of data-driven governance and homeostatic control, Visuality adapts to the crises it provokes, intensifying in response to subaltern countervisualization. Through tracing these shifts, Part One serves to situate the emergence of generative AI —not as a technological rupture, but as instrumental to the latest consolidation of authority under the imperative to secure its monopoly on visualization in the face of growing insurgency.

Visualization

Before more specifically addressing visuality and media, it's necessary to establish a frame by which we might come to understand practices and technologies of visualization as constitutive of epistemological, social, and political power. While there is no shortage of theoretical perspectives from which to approach this question, I have chosen to focus here on

bringing together a handful of canonical works that will prove useful in briefly articulating the primacy of visualization to human systems of knowledge and comprehension of the world, the effects of its simulation, and (critically) the role of particular discursive formations in determining which visualizations then come to matter and how this mattering factors into ordering the reality of social, cultural and political life.

As a starting point, I should bound the term "visualization" insofar as it will be deployed in the context of this dissertation, and its relation to practices of *representation* through media, adopting a definition extending representation's scope beyond simply the creation and manipulation of visual and auditory signs which 'stand in' for something else (people, objects, concepts, emotions, etc.) to include the semiotic prefiguration of their meaning through specific knowledge-systems or cultural circuits. Following Cassirer, who referred to humankind as *animal symbolicum* or 'symbol-making animal' (beyond the classical *animal rationale*),[2] this work thus takes for granted the supposition that practices of visualization and representation lie at the very heart of what makes us human.

According to Cassirer, what distinguishes human beings from other animals is not cognition alone, but the development of a tertiary symbolic system that mediates our engagement with the world.[3] In addition to the direct stimulus-response mechanisms shared with other species, human reactions are often subject to the interpretive delay of processing our world through abstract, representational thought. This mediation, while enabling the emergence of language, myth, religion, and culture, also distances us from the unfiltered experience of reality, having reoriented perception itself through the structures of symbolization. Cassirer goes on to explain that this adaptation has ultimately lead to humankind weaving itself a mesh of representations so dense as to effect a separation between ourselves and the physical realities we

attempt to describe, therefore living not in relation to a "world of hard facts" but the resulting impenetrable web of "fantasies and dreams" constructed by facsimile.[5] As Cassirer's characterization suggests, the benefit of such an evolution has long been questioned. While the symbolic system marks the genesis of language, art, culture, and religion, Cassirer and others have lamented the subsequent co-development of a general inability to "see or know anything except by the interposition of this artificial medium,"[6] relegating a priori knowledge of reality to the territory of philosophical pursuit, itself constrained by the logic of the symbolic system. These epistemological effects of representation were perhaps first described in Plato's "Allegory of the Cave," which described humankind's predicament as that of a prisoner, shackled and unable to experience or comprehend anything beyond their limited field of view.[7] Prevented from turning their heads, the prisoners only knowledge of reality comes from the shadows cast by a procession of "real" objects, or Forms, behind them. Naturally, the prisoners come to accept the shadows as the objects themselves, unknowingly trading reality for the experience of its double: for the prisoners, the shadow *is* reality, and suggestions to the contrary are met with much suspicion and incredulity.[8] Like Cassirer, Plato's allegory elucidates the onto-epistemological primacy of representation while at the same time grieving an unmediated experience of reality that never was.

Over time, the symbols we generate to live and act in relation to an ineffable reality have become increasingly sophisticated, developing new ways of creating and ordering representations as the visualization of a coherent, graspable reality. As the web of representation thickens to the point of solidity, its referential claim to reality dissolves in equal measure, as signs come to refer to signs in an unending loop of self-legitimizing signification. Baudrillard referred to this condition as one of simulation, wherein a complex matrix of self and inter-

referential signs (simulacra) give form to a virtual "Real" (hyperreality) eventually bearing no resemblance to the originary 'profound reality':[9] "We have all become living specimens in the spectral light of a [...] world completely cataloged and analyzed, then artificially resurrected under the auspices of the real, in a world of simulation, of the hallucination of truth."[10] I think it is here that we might find a point of articulation for beginning to describe the power inherent to visualization. Under conditions of hyperreality, where representations* no longer even feign to stand in for a profound reality, it becomes increasingly untenable to broker meaning strictly through the invocation of some transcendental signified, such as God or reality. According to Baudrillard, it is this "murderous power of images" that drove the iconoclasts, who predicted the simulacra's usurping of the "pure and intelligible Idea of God" as the guarantor of exchange between reality and its representation, and therefore strove destroy the graven image.[11] Baudrillard's theory of simulation and the idolatry of simulacra validates the aniconic concern over the invention—or reveal—of a false god, as it follows that in the absence of (or rather, blindness to) the ground of an immutable reality that exists outside its visualization, something else must emerge to underwrite the otherwise boundless play of signification, be that God or some other power.

One might see echoes of this concern over the power of representation in Heidegger's "The Age of the World Picture" where, among other things, he cites the "loss of the gods"[12] as a defining phenomenon of the modern technological age. If we draw a parallel between reality as invoked by Baudrillard and what Heidegger refers to as the totality of "what is," he characterizes

* Baudrillard marks this as a point of distinction between representation and simulation, with representation maintaining the conceit of reference, while simulation "envelops the whole edifice of representation itself as a simulacrum." (6) For the purposes of this dissertation, I adopt a use of the term representation that takes this for granted and is therefore inclusive of its later subsumption under the category of simulation.

this era as one where the epistemology of modern science as "research" has created the conditions for a conception of Being where ontological status becomes contingent on an objectification through representation. In Heidegger's formulation, "knowing" becomes conflated with a practice of research which demands that "whatever is" present itself as available for inquiry, and in turn apportions knowledge according to how readily its object is rendered calculable. In this sense, Heidegger suggests that representation is no more a mirror than a mechanism through which knowledge and Being is stabilized: only that which presents itself for objectification is granted ontological status, and truth becomes indistinguishable from the certainty that such representing provides.[13] Heidegger contends that this particular 'representing' is a distinctly modern phenomenon, an "objectifying that goes forward and masters," categorizing all of 'what is' as discrete objects of knowledge within wider systems of human comprehension.[14] What is constructed in the aggregate is therefore 'world picture,' no longer representations proffered as the image of a world, "but the world conceived and grasped" *as picture*, visualized as an internally-coherent system created and put forth by the agency of humankind.[15] This sets up a relationship between the visualization and the totality of 'what is' (which is, ostensibly, its subject) where the visualization comes to ontological dominance, with the status of 'whatever is' being contingent on its representation within it. And so it follows that in the modern age, there is a power of representation so total that the very 'being' of reality becomes conditional on its translatability into "the realm of [hu]man's knowing and of [their] having [it at their] disposal"[16] through visualization, displacing God as the ultimate guarantor of meaning and exchange. To risk overstating it: all which escapes (or is not made available through) the visualization does not 'exist.'

Stuart Hall echoes this sentiment when he argues that meaning is not inherent to an object or event but rather produced through its representation, writing that "the meaning of an event […] does not exist until it has been represented."[18] For Hall, insofar representation is ontologically and epistemologically *constitutive* of its object, it is "not *outside* the event, not *after* the event, but *within* the event itself."[19] If we accept such a conception of object/subject-hood as an effect of representation by and within particular visualizations, then we begin to see the power inherent in claiming the ability to visualize.

Clearly, this has sociopolitical implications beyond the scope of Heidegger's more strictly metaphysical account, as occlusion is the inevitable corollary of a system's taking stock of and structuring worlds in its image. Claiming the authority to determine how or whether someone or something is visualized also effects a monopoly on the means by which 'modern' societies "arrange the relations of the visible and the sayable,"[21] delineating how specific representations figure together to make up abstract concepts like history, race, politics, the nation, etc. As the authors discussed to this point have indicated, interrogating how representations (and the various technologies we employ in their production) come to matter in a sociopolitical sense then necessitates a turn from their semiotic analysis toward a consideration of the particular discourses taking shape around/between them and the roles they play in the production of meaning, knowledge and power. As Stuart Hall explains, a discursive approach shifts attention from representation more broadly to an analysis of concrete effects— "its 'politics.'"[22] This perspective asks not only how meaning is produced, but how the knowledges generated through discourse constitute 'regimes' of representation that reflect and extend particular forms of power.

The discursive approach to analyzing historically and technologically-specific regimes of visualization thus allows for an investigation of the ideological, rhetorical—and in turn, regulatory—means by which certain groups and institutions are able to bracket the otherwise boundless plays of signification, fix meanings, and cultivate a sense of naturalness or inevitability around them. Murray Edelman, writing on the political role of art, describes the power to visualize (here again, a practice that is both visual and imaginary) as not only the power to represent the world as such, but to compose the "models, scenarios, narratives and images into which audiences project and imagine politics" and in turn base their actions.[24] Sharing the constructivist notion that mental and visual images do not only represent but co-create our realities, Edelman sees politics itself as a discursive drama played out on the visualized stage of an "assumed and reported world."[25] This places visualization squarely at the root of social and political life, supplying the stable narratives and models necessary for the cultivation of meaning and the confidence to act in spite of otherwise overwhelming complexity. When the vastness of 'what is' escapes comprehension, a primary epistemological function of visualization then becomes the reduction of its multiplicities through the provision of simplified discursive models; a contracted field from which subjects may then 'freely' choose.

In *Propaganda: The Formation of Men's attitudes* (1965), his foundational work on the social psychology of propaganda, Jacques Ellul ascribes a mythopoetic function to this work, stating that the "organized myth" of total propaganda furnishes subjects with a singular, unified framework for understanding and acting in the world—leaving little room for doubt or alternative interpretation.[26] While, over the course of the twentieth century, the term 'propaganda' has developed the difficult association with a handful of violent authoritarian and totalitarian regimes, Ellul suggests that this is an unfortunate pairing. According to Ellul, as a set

of practices and techniques of representation organized according to and in support of an ideologically-inflected regime of visualization, propaganda is characteristic of modern democratic society, wherein the citizen's desire for involvement inevitably comes up against an inability (due to any combination of incidental and structural constraints) to fathom the complex dynamics they wish to engage.[27] By supplying the ready-made discursive frames which allow for the illusion of knowledge and participation, it satisfies the psychological needs of a citizenry primed to accept "a propaganda that will permit them to participate, and which hides their incapacity beneath explanations, judgments, and news."[28]

According to this framing, the provision of reductive discursive models is made to appear as much a palliative service to an overwhelmed populace as it is a strategy for epistemic capture and control by dominant cultural, economic and political interests. One result is the naturalization of an epistemic paternalism marking a strict delineation between those who claim the exclusive authority to visualize reality, and those relegated to the status of passive consumer, capable of interfacing with reality (and each other) only through the internalization of the "conceptual maps"[29] that are produced. Through the provision of discursive frames for visualizing reality, power simultaneously produces and inserts itself between world and subject.

Visuality

Having outlined visualization as a foundational practice through which meaning, knowledge, and authority are constituted, I now turn to Nicholas Mirzoeff's theorization of *visuality* to more precisely articulate its historical and political operations. Mirzoeff offers a

framework for understanding visualization not merely as symbolic or representational, but as a structured practice of governing perception. This concept anchors the periodization that follows, situating generative AI within a longer arc of Visuality's complexes—plantation, imperial, and counterinsurgent—and their entanglement with evolving media technologies and political rationalities. This framing not only positions visualization as a site of contestation, but informs the dissertation's later analysis of the regulatory, technical, and epistemic closures taking shape around generative AI.

While Mirzoeff's use of the term overlaps with definitions common to the field of visual culture, he specifically characterizes modern visuality not as an agential force, but as a "discursive practice that has material effects,"[30] chief among them being the self-legitimization of authority as an epistemological complement to its deployment of force. This authority is derived from the power to interpret and discern meaning from signs, producing new conceptual mappings that discursively order reality, invent the Other, and set the stage for normative social, cultural and political life within a society. As such, Mirzoeff's visuality is at once a medium for the production and circulation of authority and a "means for the mediation of those *subject* to that authority," invested in interpreting and representing the world in ways that make certain hierarchies and distributions of social power appear natural and self-evident.[31]

This instrumentalization of representation—naturalizing rule through aestheticization—can be traced back at least to Plato's *Republic*, wherein rulers instrumentalized falsehood and deception in the form of the *noble lie*, legitimizing political order through its aestheticization in myth. Visuality is, perhaps paradoxically, a primarily *imaginary* practice of making reality visible in both the mental and perceptual sense; synthesizing perception, information and insight in order to render the world through visualizations that are available to—and reflective of—the

authority of the visualizer.[32] A straightforward, if narrow and simplified, example of visuality at play exists in the practice of map-making, thinking of it as either the work of a military general 'visualizing' the battlefield from information gathered by subordinates, or the colonial practice of border demarcation imposing the abstract construct of nationhood onto undifferentiated physical space. In each case, through a conflation of reality with the visualizer's projection, visuality discursively "sutures authority to power and renders this association 'natural'," [33] imposing 'unreal' representations as evidence of its legitimacy.

Mirzoeff roots this formulation of visuality in the work of Scottish historian Thomas Carlyle, who in his lectures *On Heroes*[34] coined the term's use to refer to a historicising form of vision that was solely the enterprise of power. For Carlyle, visuality names the perspective of those who possessed the power and capacity to "see history as it happened," a mode of heroic 'visualizing' available to others only in retrospect, or denied altogether, as with those "minor actors of history" whose vision was seen as obscured and therefore did not constitute visuality.[35] Carlyle's framing thus cast visuality as a privileged, totalizing view-from-above: the world rendered intelligible through the sovereign gaze of those positioned to order it. This formulation exemplifies visuality as a mode of representing history from the perspective of power and therefore cast as epistemologically superior to the visualizations of those who are in turn made subject to that authority.

However, Mirzoeff contends that this is not the whole picture, claiming that the history of visuality itself must be read as the site of struggle between different modes of visualizing; ones that bolster authoritative claims to discursive power and those that oppose this monopolization of meaning. Mirzoeff characterizes this discordant way of 'seeing' (which counters and is countered by the authority of Carlyle's hero) as the claiming of the "right to look," meaning the

right of individuals and groups to have access to visuality and to be able to participate in the production, circulation and discursive framing of visualizations. To claim the right to look is to assert an equal claim to the real, or as Mirzoeff puts it: "The right to look confronts the police who say to us 'move along, there's nothing to see here.' *Only there is, and we know it and so do they*."[37] It is the recognition and assertion by history's 'minor actors' of a visualization that is a horizontal negotiation of meaning, a mutual exchange, and the claim to autonomy from the authorities that otherwise visualize and invent them as subjects. The right to look refuses segregation and "spontaneously invent[s] new forms,"[38] generating visualizations that counter the commodification and discipline of life that is characteristic of dominant visuality.

Extending this, Mirzoeff identifies two poles of visuality that are continually in contest, naming them Visuality 1 and Visuality 2. This categorization is made in reference to Dipesh Chakrabarty's identification of the two perspectives on history at work in Marx's *Theories of Surplus Value*, one which narrativizes the past as a precondition for the ascendancy of capital, and another focusing on those antecedent elements of history that do not lend themselves to the reproduction of its logics—or History 1 and History 2.[39] Following this model, Mirzoeff identifies Visuality 1 with the representation of a reality amenable to centralized leadership and control, creating a "picture of order" that sustains the disciplinary logic of capital as the authoritative "division of the sensible." In contrast, Visuality 2 is figured as the autonomous and/or collective visualization of realities which refuse subjugation, and have subsequently been denigrated by authority as "barbaric," "uncivilized," or "primitive."[40]

In contemporary contexts, this antagonistic framing persists. Increasingly, strands of Visuality 2 are also figured as misinformed, conspiratorial, or dangerous—folded into contemporary discourses of disinformation that cast countervisualization as epistemically deviant

and politically suspect. And so Mirzoeff contends that histories of power and representation can productively be read through an interaction between the totalizing thrusts of Visuality 1, which authority must adapt in order to maintain legitimacy, and the insoluble excesses of Visuality 2, which resist sublimation. Using this framework, evolutions in the form and function of Visuality 1 are understood to be driven by the development of new visualizing techniques *and* the concurrent emergence of increasingly effective strategies for claiming the right to look, often through appropriations of the same methods and technologies employed in service of authority. This drives mutations in the form of visuality through a pressure Foucault named *intensification*, with each successive stage of visuality thus having a "standard" and an adapted "intensified" form.[42]

What Foucault called the increasing *intensité* of power was a driver of its propensity to become lighter and more economical, adapting over time to transform disciplinary action into "a regular function, coextensive with society."[43] In elaborating on this intensification, Foucault describes how power adapts "not to punish less, but to punish better" through continually shifting its focus, scaling its interventions, and developing new techniques tailored to increasingly subtle and distributed targets. Through these refinements, the "art of punishing" becomes more regularized, efficient, and pervasive—decreasing in spectacle as it is optimized along expanded and differentiated circuits of application.[44] In this sense, intensification names the process by which established practices evolve to become more diffuse and internalized—less visible as force, but more total in their effects. Mirzoeff suggests that intensity similarly drives the evolution of visuality as a modality of power—one that becomes increasingly light, efficient, and evasive in response to mounting challenges to its authority—and traces this evolution through a genealogy of successive but overlapping eras or *complexes*.

As a modality of power, Mirzoeff presents visuality as functioning through a series of three complementary discursive operations. First, it classifies, imposing order through the assignment of names, categories, and definitions in an exercise of what Foucault referred to as 'the nomination of the visible.' Second, it operationalizes these classifications as a means of organization, directing the fragmentation of the social to prevent the coalescence of oppositional political subjects "such as the workers, the people, or the (decolonized) nation." Finally, it aestheticizes these arrangements, rendering them intelligible, self-evident, and thus natural or right.[46] Taken together, these operations make up what Mirzoeff names the *complex of visuality*, referring to a coordinated set of historically-specific social organizations and processes of representation affecting the "visualized deployment of bodies and a training of minds, organized so as to sustain both physical segregation between rulers and ruled, and mental compliance with those arrangements. The complex that thus emerges has volume and substance, forming a life-world that can be both visualized and inhabited."[48] As technological and cultural conditions shift, complexes of visuality adapt, intensifying until the techniques of classification, separation and aestheticization have evolved to such a degree that it constitutes the emergence of an entirely new complex. Mirzoeff's genealogy of visuality identifies three distinct permutations that have appeared in the modern era: the plantation, imperial, and military-industrial (or counterinsurgent) complexes.*

* As a note: Mirzoeff's method is perhaps best understood as a form of Foucauldian discourse analysis, linking the intensification of visuality to the broader workings of discursive power to trace how particular regimes of visualization function to naturalize domination.

Periodization Framework

In order to most effectively situate the relatively new technology of generative AI (and the correlative strategies of closure) in relation to the trajectory of media and power's co-evolution over time, I would like to propose a framework which, anchored by Mirzoeff's genealogy of visuality, corresponds this unfolding to a number of related periodizations that together might allow for an understanding of the environments within which specific media technologies come to act as brokers—and intensifiers—of power:

Periodization Map

Complex	Society	Media	Technique
Plantation	*Sovereign*	*Inscription*	*Violence*
Imperial	*Discipline*	*Index*	*Enclosure*
Counterinsurgent	*Control*	*Data*	*Modulation*

This framework draws on a longer tradition of theoretical periodization that associates specific modes of governance with their corresponding technical or media apparatuses. As others have mapped sovereign societies to mechanical machines, disciplinary societies to thermodynamic technologies, and control societies to cybernetic or computational systems,[49] this project similarly aligns each period with a characteristic media technology. In this schema: sovereign power is paired with mapping and law; disciplinary power with chemical photography

and the archive; and control with data and algorithmic media. Later, Generative AI is thus able to be situated as the media form native to the intensified counterinsurgent visuality of control societies.

It is important to note here that periodizations are, by nature, liable for varying degrees of imprecision and over-simplification, and the character of these shortcomings are nearly always convenient to the specific analyses that put them to work. Arguments making use of composite periodizations, like the one above, should be especially mindful of the fact that periodization theory is, at best, an "analytical mindgame, yet one that breathes life into the structural analyses offered to explain certain tectonic shifts in the foundations of social and political life"[50]—and this one suits mine. But that is not to suggest that the correlations I've indicated above are arbitrary. While the progression between periods does not always map neatly over one another (indeed, they shift slowly and out-of-sync, overlapping one another and themselves, the prior sometimes never fully receding as its successor comes to dominance) it is precisely within this incongruity that we are able to see the drivers and processes of power's intensification, as shifts in one area exert pressure on the others, which ultimately adapt.

In the following pages I will discuss the dynamics of each period, highlighting the role of media technologies in the production and exercise of knowledge as power, and to what extent they act as drivers of its intensification. A comprehensive account of these histories is well beyond the scope of this dissertation, but an overview is necessary to sketch the continuities of real and discursive power into which generative AI are emerging as an intensification of counterinsurgent control. With this is mind, I will be paying more attention to those conditions specific to the latter two complexes, though it bears noting that new techniques of visualization

do not supplant their priors, but rather augment or extend them through intensification, with all prior apparatuses simultaneously at work in the complexes that follow.

Plantation/Sovereign/Inscription

Lasting from the seventeenth to around the late nineteenth century, the first modern complex of visuality identified by Mirzoeff was that of plantation slavery and the particular orderings of reality that sustained it, "summarized as mapping, natural history, and the force of law."[51] Represented in the enactment of the Barbados Slave Code and Louis XIV's *Code Noir* in the mid-to-late seventeenth century, the legal work of categorizing and separating human beings as property emerged in Europe contemporaneously with the taxonomic "order of things"[52] that was produced as an effect of recent developments in the discourse on the human and natural sciences, thus forming grounds for "the possibility of representing space and species in distinct but exchangeable form."[53] These developments made possible an accounting of natural history whereby a "certain set of people were classified as commodifiable" and the extraction of their unwaged labor from 'cultivated land' (another colonial projection) made legally commensurate with that of other 'natural' resources like mining, forestry and agriculture.[54]

Already in process of being aestheticized through force of law, natural history's classification and separation of people according to newly defined categories of race was further concretized in the colonies through the relatively recent technology of mapping, which identified/inscribed these separations on the earth through the cartographic visualization of land as either 'free' space or 'slave' space. Mirzoeff emphasizes mapping as an early example of a

media which acts as a "technical prothesis" enabling visuality's manifestation as force, noting that, like other media, the map performs precisely the fusion of perceptual and conceptual vision that would later be claimed as visuality.[55] Such novel manipulations of space through the visual technology of mapping enabled the imaginary projection of space and people as either 'free' or 'slave,' 'cultivated' or 'empty'; divisions that were then enacted and maintained through the mechanism of oversight, a centralized regime of surveillance and spectacular violence whereby "even the abstraction of the law was made visible and performative."[56]

This colonial "order of things" was also aestheticized through the publishing of many manuals for the work of colonization and planting. In these manuals, practical instruction was accompanied by many tables, maps and engravings, schematic and pictorial representations visualizing "both the process and the power that sustained" the well-run plantation.[57] Cataloging the hierarchical divisions of personhood and labor that would come to characterize the new world plantation, these manuals (together with the aestheticizations of 'natural history' and the force of law) served to naturalize domination through a discursive conflation of the practice of oversight with the authority of the distant European sovereign. Codified as the surrogate of sovereign power and its feigned monopolies of force and authority, the solitary overseer (depicted at the center or top of diagrammatic inscriptions, surveilling a workforce that vastly outnumbered him) thus embodied the visualized techniques of this authority, and was in turn repeatedly represented as the conduit for a gaze which could alone compel "unwaged labor to generate profit from the land."[58]

Taken together, the colonial techniques of mapping, natural history, and the force of law point to one of the earliest formations of visuality as a discursive practice leveraging emergent technologies and methods of representation in ways which make domination appear natural and

therefore legitimate. By the same token, the various challenges that would emerge in response to the visuality of oversight—through increasing access to representational apparatuses and subaltern claims to self-representation—illustrate how changing conditions exert pressures that eventually lead to visuality's intensification. In the case of the Plantation Complex, Mirzoeff registers these challenges under the banners of Revolutionary and Abolitionist realism, two successive but overlapping countervisualities that aestheticized emancipation through the figure of the revolutionary hero (who countered the sovereign authority embodied in the overseer with a hero self-authorized by the people); legal dissensus and the performative claim to rights whereby the enslaved "renamed themselves as the people, classified themselves as rights-holders, and aestheticized their transformation in the person of the [revolutionary] Hero";[59] and the figurative and literal remapping of colonial space, notably by the Maroons (escaped slaves) who's persistent attacks on the coastal plantations pressured the English and Dutch into legal arrangements mapping a "permanent zone of exclusion in which slavery was not permitted," appropriating two means of aestheticization (mapping and force of law) for anti-slavery ends.[60]

As the enslaved (along with still marginalized allies like Jews and free people of color) together asserted the right to look, these pressures were increasingly destabilizing the plantation economy, which had in turn become legally and practically cumbersome. In the mid-to-late eighteenth century, the plantation complex underwent a succession of intensifications in response to the ensuing breakdown of the supervisory regime of oversight, automating large swaths of production and adopting new techniques of statistical management,[61] signaling the arrival of a modernism that would be characteristic of the emerging period.

Imperial/Discipline/Index

As oversight's rigid divisions of space, personhood and labor attenuated under the intense pressure of challenges to its authority, Visuality likewise adapted to the changing conditions. As the legal* reality of 'free' and 'slave' identities fell out over time, there emerged an intensified visuality which would authorize power along a new, more nimble and expansive hierarchy of the 'civilized' and 'primitive'. In this iteration, imperial visuality restructured its classificatory logic around a spatiotemporal ordering of history. Rather than authorizing domination through legal status, it positioned colonized, racialized, and otherwise deviant populations as coexisting in time with their colonizers, but visualized them as *historically* prior. In this schema, 'civilized' is imagined as the historical vanguard endowed with the authority to modernize, govern, or separate and contain the 'primitive.'[62]

Trading the legal categories of free/slave for a system of classification based on the cultural/historical hierarchies of the civilized/primitive (with civilization embodied by an ascendant European modernism), imperial visuality not only served as continued justification for domination and expansion in the colonies but also marked a point of articulation for the extension of centralized imperial power into the metropole. The new complex of visuality thus effected a separation that was operative across and between geographic and political borders, constructing a cultural discourse whereby the mentally ill, criminals, "metropolitan working class, [and] inhabitants of the "South" within Europe […] were considered closer to the actually existing "primitives" of the South Pacific than to the imperial elite,"[63] an attitude reflected in the

* This is, of course, in contrast to the social and political realities of slavery, which persisted and do persist.

not-uncommon 19th century practice of exiling convicts and the mentally ill alike to the colonies.*

The re-invented hierarchies of imperial visuality would be concretized by new techniques of classification following from the rising tide of statistical management, the subsequent invention of the concept of a social and biological 'norm' (*l'homme moyen*),[64] and echoes of slavery along what Frederick Douglass,[65] W.E.B. Du Bois[66] and others named 'the color line,' marking the imagined divide between lighter and darker-skinned individuals which was, unsurprisingly, mapped onto the new dichotomy of normal/deviant.[67] Framed as part of the civilizing project, this discursive work of classifying, separating and visualizing the social body along "scientific" lines thus complemented and legitimized the newly decentralized techniques of imperial power manifesting in the form of disciplinary institutions such as the prison, asylum, factory and school. Such a project would benefit from a media with a similar air of scientism, and the recently invented technology of photography fit the bill.

While people were experimenting with creating images using light-sensitive materials as early as 1717,[68] it wasn't until well into the nineteenth century that techniques were developed which allowed these reactions to be permanently fixed.† The invention of a practical and commercially viable photographic process marked one of the greatest paradigmatic shifts in media history to that point. Created through an ostensibly 'automatic' and scientific process, and therefore by-passing "direct involvement of the mental states of the image-maker,"[69]

---

* Typified by the French Relegation Law of 1885, which relocated recidivist criminals to penal colonies in Guyana and New Caledonia: Jean-Lucien Sanchez, "The Relegation of Recidivists in French Guiana in the Nineteenth and Twentieth Centuries," in *Global Convict Labour* (Brill, 2015), 222–48.

† In 1822, Nicéphore Niépce made the first permanent photographic image through a process he called heliography. Willfried Baatz, *Photography* (New York: Barron's, 1997), 16.

photography was seen as having an indexical relationship to reality that handmade inscriptions (no matter how expertly rendered) could not: “a photograph is not only an image (as a painting is an image), an interpretation of the real; it is also a trace, something directly stenciled off the real, like a footprint or a death mask”[70]—the first defensibly objective medium.

In *Discipline and Punish* Foucault places the emergence of the modern disciplinary institution (exemplified by the prison) in the 1840’s,[71] a date conspicuously contemporaneous to the appearance of the new technology of photography, which quickly found its place as the primary instrument of identification and surveillance. Given its aura of transparency and scientific precision, photography effectively standardized the creation of visual representations in line with the disciplinary emphasis on the documentation, categorization and management of individuals according to specific statistical distributions, effectively transforming the body into an object of knowledge and control. As such, photography provided material for a growing bureaucracy tasked with defining and cultivating a normative conception of the human that would serve to not only further justify the wholesale domination of oversea populations, but similarly stratify the domestic social body to extend colonial biopower into the heart of the European capitol.

Allan Sekula wrote that this appropriation of photography by the repressive logic of the disciplinary apparatus constituted the “lower limit or ‘zero degree’”[72] of an instrumental realism that would come to define the social character of the medium in the nineteenth and twentieth centuries. Sekula identifies this logic at work in the projects of Alphonse Bertillon and Francis Galton, which, though methodologically opposed, shared a common positivist impulse to define and typologize deviance through the convergence of visual representation and statistical reasoning. Bertillon, a Paris police official, developed a two-part system of criminal

identification merging photographic portraiture, anthropometric measurement, and written annotation into individual records (*fiches*), which were then organized into statistically indexed archives—effectively embedding and locating their subject within a larger social order. Around the same time, Galton used photographic composites in an attempt to conjure an "optical apparition of the criminal type," abstracting averaged physical features into a visual ideal of deviance that was, as Sekula notes, "empirically nonexistent."[73] Despite their differences (Bertillon's system enabled the literal and figurative apprehension of recidivist criminals by state apparatuses, whereas Galton's composites ultimately provided a quasi-scientific justification for the goals of the international eugenics movement) both projects enlisted photography as a visual technology of imperial domination, fixing the body as evidence of its real or imagined transgression.

Sekula contended that the influence of this repressive logic on the early discourse around photography's social meaning was such that even the honorific functions of the medium could not override its social and political influence. As vernacular deployments effectively "welded [its] honorific and repressive functions together," even the common photographic portrait, he suggested, functioned as a tool of panoptic self-discipline.[74] As a visual artifact now situated within the broader social hierarchy, the portrait's private meaning became laden with both the more public, upward gaze toward authority, and the downward glance one's 'inferiors.' For the layperson, then, even self-representation became enmeshed with the internalized disciplinary gaze of imperial authority.

Within this emergent regime of visualization, the epistemic value of privately held albums and institutional archives was not inherent, but determined through a discursive politics of representation structured around what John Tagg described as "the division between the

power and privilege of *producing* and *possessing* and the burden of *being* meaning."[75] The result was a photographic 'regime of truth' which inscribed meaning and agency unevenly across subjects, generating and legitimating the boundaries around which representations were allowed to signify. Originating in Foucault, the concept of a "regime of truth"[76] refers to the systems of knowledge production and dissemination that are operative within a particular social formation and time period. By linking the epistemological notion of 'truth' to the political 'regime,' the concept rejects essentialist conceptions of truth as a universal with a fixed relation to reality. Instead, it understands truth as a discursive formation produced by social and political forces with vested interests in determining which representations are considered authoritative, how true and false statements are distinguished, by whom, and under what conditions.[77]

In *The Burden of Representation* John Tagg also used the instrumentalization of photography in police and legal work to discuss how photographic discourse in nineteenth century Europe and North America developed according to a *realist* regime of truth which put in place strict rhetorical and technical conditions that must be met in order for photographs to be invested with evidentiary value. Quoting at length from reference books on photographic evidence and practical police manuals, Tagg highlights a recurring emphasis on themes of "sharpness and frontality, exhaustive description and true representation, the outlawing of exaggerated effects and any overt kind of manipulation, standardised processing and careful presentation, and the expertise and professional status of the photographer" as the aesthetic and procedural criteria necessary to establish photographs as objective records of "legal quality."[78]

As a side-effect of standardization, discourse around the epistemological value of photographic representation developed in such a way as to effectively de-authorize images produced by individuals and collectives that were not affiliated with or sanctioned by the

authority of 'objective' institutional entities. Tagg contends that within this regime of truth, realism functions through a social practice whereby credibility is derived not from an indexical relation to reality but through its intertextual reinforcement across 'sanctioned' representations: a "mutuality which summons up the power of the real: a reality of the intertext beyond which there is *no-sense*."[79] The realist mode thus depended on repetition and citation to instantiate its reality, which then functions as a disciplinary norm in place of any onto- or epistemological given.

In a later paragraph more or less describing Mirzoeff's visuality, Tagg stresses the point that, despite claims to immutability, the realist process of signification was a historically-specific production contemporaneous to the development of disciplinary society in the eighteenth and nineteenth centuries. By establishing this convention across the emergent complex of state institutions, the ruling class was able to aestheticize the separations central to the administration of imperial visuality, fabricating an identity between the realist signifier and the immutable signified.[81] Under these conditions, the epistemological value of representations (photographic and otherwise) would then be evaluated according to a strict positionality "whose reference point was the *dominant discourse* which appeared to have its point of origin in [reality]."[82] As such, the realist convention inaugurated an epistemological condition whereby subaltern subjects could be denied the right to look despite rapidly democratizing access to the 'objective' technology of photography.

As the photograph's authority is derived less from its intrinsic indexicality than the privileged conventions and institutional contexts in which it operates, Tagg writes that the "power to bestow authority and privilege on photographic representations is [simultaneously] not given to other apparatuses within the same social formation," such as popular, artistic, and activist photography.[83] So there emerges alongside photography a regime of truth centered on

the installation and propping up of the false dichotomy between who *produces* meaning through representation and who representations produce *as* meaning, turning people "into objects that can be symbolically possessed"[85] and manipulated within bureaucratic systems of classification. In the terms of imperial visuality, this split both mirrored and was constitutive of the complex's binary classifications; coding the colonized, working class, and mentally-ill alike as objects of the disciplinary gaze, "forced to emit signs, yet cut off from command of meaning[…] incapable of speaking, acting or organising for themselves."[86]

Of course, according to Mirzoeff, challenges to the authority of imperial visuality would still emerge in the form of proletarian and antifascist neorealisms concerned with reclaiming the right to look for a populace made subject to imperial categorization and control. Where antifascist neorealism (fascist aestheticization being the intensified modality of imperial visuality)[87] largely visualized this claimed autonomy through the production of art and cultural theory highlighting the "contradictory nature of a conflicted reality held in place by the operations of the police" and imagining new realities in its stead,[88] proletarian countervisuality took the general strike as its primary tactic against the hierarchies of imperial visuality. Countering the atomization—or disaggregation—of the social body by the disciplinary gaze, the general strike was an exercise in aestheticization. Marking the moment when the strength of the populace is made "visible to itself," Mirzoeff contends that the mass refusal of the general strike was to be "understood not simply as a work stoppage but as the right to look: to see and be seen *by each other* within a historical moment and to know how things stood."[89] Against the separations effected by their classification as imperial subjects (in both the colony and the metropole), workers thus visualized their aggregation as *the people*. Under the cultural weight of antifascist neorealism and the masses self-realization in the image of the general strike, Visuality

would again intensify in kind, this time along two related fronts: the rise of non-representational data as the predominant medium for the production of social knowledge and control, and the inauguration of the era of mass persuasion following the publication of Gustave Le Bon's *The Psychology of Crowds* in 1895.

While data played a large part in pre-photographic visualization (thinking back to the graphs, tables and practical manuals which aestheticized the plantation complex's hierarchies of personhood and production) the roots of data's ascendancy as the media of modern knowledge production and control can be found in the large-scale photographic projects of the nineteenth century. As the technology of photography became more widely accessible and an increasing range of disciplines put it to work, there emerged the problem of over-production. Photography's early show of promise as the infallible medium of a frothing scientific empiricism had begun to fade in inverse correlation to the increasing glut of images. While the "lingering prestige of optical empiricism was sufficiently strong to ensure that the terrain of the photographable was still regarded as roughly congruent with that of knowledge in general,"[90] it quickly became clear that without deference to a rigorous organizational scheme the task of generating meaning and extracting useful knowledge from the medium would become increasingly impossible. Sekula observes that, by demonstrating the utility of compressing visual information into an alphanumeric shorthand, Bertillon's police work offered a model of data rationalization that other fields later sought to emulate, and between 1880 and 1910, the archive would emerge as a preeminent site of knowledge production.[91]

Bertillon devised a system of classification whereby the 'aesthetically neutral' photograph constituted both the object and means of data collection about its subject in the form of eleven bodily measurements. Recorded as a standardized series of numbers and set against a

statistical norm, this ‘anthropometrical signalment’ formed the backbone of an archival system from which information about individuals could be quickly and easily retrieved.[93] Despite his presumed intent to preserve the primacy of the index, by implementing the photographic archive as a “medium from which exact mathematical data could be extracted”[94] and seemingly fulfilling the dream of universal translatability, Bertillon’s system signaled the beginning of a paradigmatic shift wherein photographic representation would be usurped by abstract, statistical procedures in the search for meaning and a surveillant mode of social control more resistant to countervisualization.

To a similar degree, the work of Francis Galton might also be read as a precursor to modern data science for its emphasis on the extraction of meaning from large-scale data collections through statistical correlation and pattern recognition. As the technology of photography democratized and the production and sharing of images became commonplace, the media itself would became synonymous with surveillance, with even the most innocuous productions ultimately feeding the large scale repositories of data which would later fuel the statistical and discursive operations of counterinsurgent visuality. This intensification marked a response to the failure of imperial visuality to sustain the legitimacy of its classificatory and political order under the growing pressures of decolonial countervisualization—a crisis in which the aesthetic division between the “civilized” and the “primitive” could no longer be maintained.[95]

Counterinsurgent/Control/Data

Counterinsurgency

Along a period spanning from about the 1920's until the end of the second world war, the imperial complex—pushed along by the pressures of countervisual resistance and rapid technological change—would again gradually share space with and morph into a newly-intensified paradigm of visuality. In a turn to what Mirzoeff dubbed the "military-industrial"* complex, the visuality of global counterinsurgency differed in that its power would no longer hinge on the institutionalized punishment and reform of subjects deemed to be non-conforming with the characteristic binaries of the imperial complex. Instead, counterinsurgency operates on the "technical *management* of neo-imperial dominions"[96] through an instrumentalization of information and culture as the 'operational codes' by which beliefs and behaviours are conditioned. In contrast to the onerous task of producing disciplined, docile bodies, counterinsurgent visuality intensifies by unburdening itself of the expectation of individual reform, focusing instead on the behavioral science of modulatory population control; managing the terrain of culture and information so as to legitimize domination, render populations governable, and produce consent at a macro-level.[97] As a strategy of power, counterinsurgency finds its roots in military doctrine, combining lethal force, spatial control, and ideological reproduction under the formula 'clear, hold, and build,' referring to the complementary operations of removing insurgents through violence, maintaining that separation by physical means, and establishing neoliberal governance in the secured space.[98]

* While Mirzoeff uses "military-industrial" to refer to the current complex of visuality, I find "counterinsurgent" more effectively isolates the period's intensified regimes of classification and separation from their martial origin, and therefore better suited to my specific analysis.

As laid out by the United States Army in a 2006 manual on counterinsurgency operations, COIN doctrine is comprised of "offensive, defensive, and stability operations" that are both mobile and context-specific, requiring a balance of standard combat operations with humanitarian, political, economic and security measures "more often associated with nonmilitary agencies."[100] Expanding the strategic goals of conventional military intervention to include the amorphous objective of winning 'hearts and minds,' COIN forefronts the domain of *culture* as a new theatre of operations in the battle against the irregular forces typical of post-modern warfare. Considering that culture conditions the ideology and action of populations through the formation of norms around "what to do and not to do, how to do or not to do it, and whom to do it with or not to do it with," its strategic importance stems from the fact that it also establishes the circumstances and processes by which those norms might *change*.[101] And so it follows that counterinsurgency requires high degrees of understanding, interaction, and intervention in the cultures within which it operates in order to effectively contribute to the expansion of neo-imperial influence.

Continuing with the United States as an example, military counterinsurgencies from the last century include the Vietnam, Afghan and Iraq wars (notably, each conflict was terminated before the invading forces could claim victory by any viable metric of success—more on this open-endedness later). While each included territorial control among their early strategic objectives, these conflicts crucially took place within the context of global ideological struggle (against the spectres of communism and global Jihad, respectively) and were therefore ultimately defined by protracted COIN operations aimed at 'excising' the insurgent from host populations so that they could then be made culturally receptive to the governance of a western-style client state. Referencing a section of the Counterinsurgent's Field Manual likening insurgency to

cancer and its removal to a life-saving procedure, Mirzoeff notes that, at the point of intensification, "counterinsurgency now actively imagines itself as a medical practice," extending a metaphor that, while imperfect, underscores the perceived urgency and existential threat of insurgency, justifying a necropolitics wherein the insurgent is figured as metastatic force to be eliminated so that the host society can live.[103] This imagining of counterinsurgency as biopolitics thus invests the neo-imperial force with "a power to foster life or disallow it to the point of death"[105] according to an ideology whereby 'health' and 'living' are made commensurate with liberal democracy and economic integration with global neoliberal capitalism.

While the counterinsurgent complex marks a point of significant change in the genealogy of visuality, the enduring legacies of plantation and imperial discipline remain visible not only in the initial violence of counterinsurgency (clearing), but also its instrumentalization of the technologies of mapping, photography, and archiving for the purposes of visualizing populations as data to be statistically managed (holding). During the Vietnam war, the US Army developed the Hamlet Evaluation System (HES), a program of physically isolating South Vietnamese hamlets (roughly 12,000 in total) from the surrounding countryside with the goal of eradicating Viet Cong (VC) sympathies among the population. As such, the system laid out protocols for the collection and assessment of data on the state of "pacification progress" in each location; a decidedly qualitative metric that was independent of (and often in inverse relation to) the status of US Army territorial control.[106] Through statistical analysis of data collected from across South Vietnam, the HES was an attempt to categorize hamlets on an scale from ideologically 'secure,' 'contested,' to 'VC,' thus mapping the progress of the counterinsurgency effort and providing information from which COIN strategy could be adjusted in response to the specific

conditions of each location. More recently the US Army has similarly deployed biometric tools like the Handheld Inter-agency Identity Detection Equipment (HIIDE)—which collected and archived biometric data on individuals through photographs, iris scans, and fingerprints—to "assist in identifying detainees in the war zones in Iraq and Afghanistan, and in differentiating friendly individuals from insurgents and terrorists."[107] Having amassed an extensive archive linking individual biometric data to records of affiliation with insurgent groups, the HIIDE and technologies like it visualize populations as a body to be scanned and cleared of aberration. In military counterinsurgency we plainly see the development and deployment of technological surveillance in the context of a visuality oriented toward classifying, separating and expunging the insurgent.

As a cultural and political mode of conflict that is without borders, counterinsurgency similarly blurs the distinction between foreign and domestic space. In the late-twentieth and early-twenty-first century, globalization, transnational telecommunications, and the spectre of international terror and foreign influence gave rise to a conflict fought as much in New York or Toronto as it is in Baghdad or Kandahar. What's created is something of a "nomadic border that can be instantiated whenever [the state apparatus or its surrogates] looks at a person" suspected of insurgency, either at home or abroad.[108] The resulting adaptations of domestic culture and politics in the West to the regime of counterinsurgency are perhaps most visible in the evergreen bogeymen of Socialism, religious radicalism, and the now widespread practice of bestowing even non-violent cultural opponents or political dissenters with the ignoble label of 'disinformation actor' or 'domestic terrorist.' Perhaps, then, the cancer metaphor was more complete than Mirzoeff suggests. At the level of population control, the dogma of counterinsurgency inevitably leads to obfuscations of the previously clear distinction between

civilian and combatant; by visualizing every individual as a potential vector of insurgency, the complex naturalizes its imposition of indiscriminate surveillance and control to the detriment of core democratic values—a radiation treatment of sorts.

Control

Having effectively been 'cleared' by the centuries-long onslaught of sovereign and disciplinary power, authority over colonized and domestic metropolitan populations alike could now be 'held' through a system of management based on the imposition of various affordances, constraints and manipulations. No longer founded on either the centralized authority of the sovereign or the decentralized power of the disciplinary apparatus, counterinsurgent visuality emerges as the discursive formation native and complementary to what Deleuze identified as the new *societies of control*. Whereas the institutional enclosures of Foucault's disciplinary society functioned as *molds* (castings of the productive, docile subject into which the individual and collective body was 'pressed' with steady force) Deleuze's control (which he expanded from Burroughs)[109] exercises an intensified power through targeted, ongoing, and often imperceptible processes of *modulation*.[110] Alex Williams explains that, contrary to the comparatively coarse operations of discipline, control society is cybernetic, functioning through a complex of homeostatic feedback loops wherein deviations from a particular goal or system state are counteracted through the imposition of modulatory stimuli.[111]

Recognizing the societal trend toward financialization—wherein the ambiguous 'market' unseated the institution as society's primary site of economic, social, and cultural production—

Deleuze would declare that to be subject to authority in the emergent societies of control is to no longer be *enclosed* (by factories, schools, prisons, etc.), but to be in *debt.*[112] Despite the relatively head-less nature of this new control, authority nonetheless reserves the ability to set the parameters of homeostatic regulation and therefore guarantee this debt through the maintenance of perpetual asymmetries in knowledge, expertise, and capital. The resulting mechanisms of control thus hinge on an instrumentalization of cultivated debt and desire as the primary drivers of social self-regulation. In the early-to-mid twentieth century this manifest in the complementary apparatuses of mass media propaganda,[113] corporate advertising, and the culture industry,[114] which together visualized a contracted field of knowledge and possibility wherein the breadth of individual actions became reduced to a set of pre-determined choices "*already made* in the sphere of production".[115] Within this narrowed field, propaganda directs behaviour not only via the control of information but through the constitution of subjectivity itself, "[taking] hold of the entire person" through the imposition of a totalizing framework of intuitive knowledge.[116]

Jacques Ellul characterized such propaganda as a "modern scientific system" that requires constant evaluation and adjustment in order to best achieve its desired effects.[117] Operating as a large-scale homeostatic mechanism, it realizes these effects through a near-total capture of the means of knowledge production and dissemination within technologically-advanced society. Once intensified as total propaganda, control shifts its objective from the cultivation of society-wide adherence to a particular social or political doctrine to the targeted production of various states of *orthopraxy*: a condition of habitual reflexive action that "in itself, and not because of the value judgments of the person who is acting, leads directly to a goal."[118] In line with the shift from discipline, as control evolved this 'goal' would similarly move away

from the production of a homogeneous social body toward the management and stabilization of a heterogeneity from which value could reliably be extracted. Mirroring the shift to what Foucault would call biopolitics, authority's interest in the soul and body was soon replaced by control's desire to apprehend 'life itself.'

Internet & Protocol

Over the course of the twentieth century, the practices of homeostatic modulation perfected through the apparatuses of print, radio and television would lay the groundwork for the further intensification of a counterinsurgent visuality based on control over 'life itself,' with the development and spread of advanced information and communication technologies, most notably the internet and its later maturation around the platform model.

With the invention of the internet, the technological foundations were laid for the widespread operationalization of control mechanisms as a tool of social management. By reducing the technical and economic burdens associated with large-scale yet targeted information gathering and dissemination, the creation of decentralized networks of computers enabled the "flexible and fluid employment of homeostats throughout society" at scales and speeds that were impossible within the context of centralized mass media.[119] As a defining characteristic of the internet, decentralization thus represented a hurdle for authority insofar as it displaced the well-established paradigm of one-way communication, even as it extended its capacity for modulation by embedding control directly into the technologies that make communication possible.

In *Protocol: How Control Exists after Decentralization*, Alexander Galloway elaborates on this seeming contradiction, noting that while digital networks become increasingly distributed and omnidirectional, they remain fundamentally hegemonic insofar as they're governed by a "negotiated dominance of certain flows over other flows."[120] According to Galloway, control thus becomes an effect of the various technical procedures—or protocols—that handle the flow of information across decentralized networks. By establishing and enforcing the "conventional rules that *govern the set of possible behavior patterns* within a heterogeneous system," Galloway presents protocol as central to the internet's functioning as a "new apparatus of control" which has come to dominance in the twenty-first century.[121] The very existence of protocol thus challenges the increasingly outmoded notion of the internet as a fundamentally democratic technology enabling free and open communication, with the appearance of unrestricted exchange being the aesthetic product of a strictly-defined set of technical procedures which "[enables] such properties to emerge."[122]

When all things are functioning as designed, there is no communication on the internet without control. Galloway traces the foundation of this control to the contradictory functions of two protocols central to the modern internet: TCP/IP (Transmission Control Protocol/Internet Protocol), which enables unencumbered peer-to-peer connection between all computers on a given network; and the DNS (Domain Name System), which regulates that connectivity. As a database correlating network names to their actual locations, the DNS protocol is essential to nearly all internet communication. However, despite the appearance of decentralization, the DNS system operates through a strictly hierarchical, inverted-tree structure in which central servers wield control over the connective possibilities of the nodes below them. Due to this descending order of governance, anchored by a small handful of 'root' servers, the DNS architecture

effectively centralizes power over which nodes can access (or be accessed by) other nodes on the same network, a reality which, Galloway notes, "should shatter our image of the Internet as a vast, uncontrollable meshwork."[123]

While this is clearly a gross simplification of how protocol works,[125] it provides us a useful framework for understanding protocol as a mechanism of counterinsurgent visuality and control. Insofar as it purports to both construct and regulate access to an exhaustive index of the internet, the DNS and other protocols assume the authority to mandate what connections are permissible on the network, and in turn limit what realities individual users are able to visualize with the data they encapsulate. Galloway reminds us that "the limits of a protocological system and the limits of *possibility* within that system are synonymous," with access to visualization being contingent on adherence to pre-established codes of meaning:[126] "DNS is not simply a translation language, *it is language*. It governs meaning by mandating that anything meaningful must register and appear somewhere in its system. This is the nature of protocol."[127] In the context of counterinsurgency, this control can be thought of as functioning through the creation and management of degrees of *in*visibility, utilizing protocol to disrupt communication and restrict access to subversive actions or information without the need to undergo the resource intensive work of removal and discipline. Protocological analysis like the one Galloway provides reveals the internet as a technology of contraction. Torn between liberation and standardization, it contains "both the ability to imagine an unalienated social life and a window into the dystopian realities of that life."[128] At the time of writing, the reality of networked social life must be understood in relation to two related trends: platformization and algorithmic regulation.

Platforms, Algorithms & Machine Learning

Beginning in the late nineteen-nineties with the appearance of e-commerce sites like Craigslist and eBay and maturing over the course of the aughts with the rise of Amazon, Twitter and Facebook and others, the platform has come to be the predominant model for social and economic activity online. At their core, platforms provide basic digital infrastructures enabling users to communicate or engage in commerce, while also serving as a base upon which they're able to create and serve their own products and services.[129] Providing easy-to-use interfaces and free access to markets and audiences, these sites benefit from a network effect whereby their usefulness increases in direct correlation to the number of users on the platform. As a result, the model has captured a vast majority of social, economic and cultural life online, generating a new political economy of information wherein power stems not from the production and exchange of new products and information (which is now the domain of users) but from the appropriation and exploitation of statistical information produced in the course of social and economic activity.[130]

By inserting itself between users as the literal "ground upon which their activities occur,"[131] the platform is uniquely positioned to indiscriminately record this data and to operationalize it in service of the complementary goals of profit generation and counterinsurgent control. The success of platforms in attracting and retaining large user-bases is itself illustrative of a balance between restriction and enablement that is necessary to their operation as mechanisms for extracting value from certain human behaviours while at the same time modulating, transforming, or precluding others. Williams notes that this balancing of openness and constraint is precisely what has allowed platforms to emerge as potent vectors of modulatory control: enabling a wide-range of user interactions and behaviours while simultaneously

modulating those same activities through the iterative refinement of platform affordances.[132] Platforms' framing of contingency is the product of two related features: their design around a core architecture governing the limits and form of interaction possibilities; and the management/modulation of content visibility and user behaviour by machine learning algorithms. While originally developed as complex (yet relatively volatile) adaptive systems reliant on the openness and unpredictability of user interaction, platforms have since matured into powerful (often unquestioned) mechanisms of control that, "given the appropriate socio-technical resources, can be effectively operationalized, and hence can emerge as a political technology."[134]

In the context of counterinsurgent visuality and control, this power stems from platforms' de facto authority to visualize the reality of our digital lives, surfacing select content and interactions from the glut of user-generated information through the algorithmic operations of search, trending and personalized recommendation. Leveraging machine learning (empowered by the large-scale data collection made possible through platform permissiveness) to analyze, predict and modulate network behaviour, algorithmic control functions to amplify information and actions amenable to the objectives of authority, while simultaneously working to "suppress irrelevant variations."[136] As such, these algorithms share a fundamental trait with both visuality proper and the protocols that make platforms and network connection possible, as both (though they are so intertwined that to address them as distinct risks misapprehending them) are as concerned with *dis*connection and *non*-communication as they are with connectivity. While their primary objective is to establish connection and direct flows of information, it is in the moment of disconnection that platform algorithms "most forcefully display [their] political character."[137] Once operationalized as counterinsurgency, platform infrastructures democratize the means of

communication all the while securing *visuality* for dominant interests, precisely modulating behaviour and visibility in order to keep insurgent possibilities out-of-sight or otherwise contained.[138]

As much as they function as regulatory technologies of counterinsurgency, machine learning algorithms are also powerful agents for visualizing realities from data which reflect and further naturalize the entrenched discursive and aesthetic conventions borne out of plantation and imperial visuality. Problems of algorithmic bias—as addressed by scholars like Safiya Noble (who challenges the notion of technological neutrality, particularly as relates to search and recommendation algorithms that routinely amplify existing racial and ethnic biases),[139] Ruha Benjamin (who relates machine learning to the discursive codes of racism insofar as both are *predictive models* "powered by haphazard data gathering and spurious correlations, reinforced by institutional inequities, and polluted by confirmation bias")[140] and others[141]—illustrate how, in both form and content, the classificatory and protocological operations at the heart of machine learning tend to mirror the discursive operations of classification, separation and aestheticization central to the dominant orderings of reality that Mirzoeff refers to as Visuality 1. In spite of a continuously renewed rhetoric of objectivity and progress, the case remains that machine learning and other artificial intelligence "could not exist without data produced through histories of exclusion and discrimination."[142] It follows then that visuality is not external to (and therefore eradicable from) technical design, but *co-constitutive* with the mathematics that make algorithms 'work.' Therefore the mis-categorizations and mis-representations we refer to as bias are not—in a sense—mis-anything at all, but the result of predictive algorithms functioning as designed, generating visualizations in accordance with discursive separations effectively naturalized over five centuries of modern visuality. Applied in the context of social networking, news, and e-

commerce platforms, algorithms thus enact processes of signification according to an ingrained desire to *distinguish*, performing what Bernhard Reider calls "interested" readings of reality—not in the sense of being biased, per se, but in the sense of optimizing the function to distinguish above all else. Emphasizing his point with a quote from Lyotard, he explains that the epistemic processes put into play by algorithms are explicitly designed around success rather than accuracy: "The goal is no longer truth, but performativity—that is, the best possible input/output equation."[143]

In the context of counterinsurgent visuality, the performativity of platform algorithms refers to their probabilistic management and control of heterogeneity, and "the interested readings these techniques perform fit into a system that attempts to extract monetary value from the exploitation of increasingly infinitesimal differences."[145] In his 1975 essay "The Limits of Control," William Burroughs writes that while societies of control try to "make control as tight as possible, […] if they succeeded completely, there would be nothing left to control."[146] Control requires opposition in order to totalize its authority, and as such extends the doctrine of counterinsurgency to imbue everyday life with the perpetual risk of upheaval, carefully balanced so as to avoid "a showdown where all-out force would be necessary."[147] As the discursive framework for a system of control oriented toward managing and exploiting dissension, counterinsurgency centers culture as "the means, location, and object of warfare,"[148] and in turn gives up the pretense to visuality's discursive *aestheticization* of a reality in legitimation of its separations, instead falling back on a nihilistic reframing of classification and separation as *ends in themselves*,[*] with algorithmic media playing a central role.

[*] In *Democracy and Other Neoliberal Fantasies: Communicative Capitalism and Left Politics* (2009) Jodi Dean nods to the reliance of liberal and "left" politics on a powerful opposition as the pretext for its perpetual

Countervisuality and the Right to Look

Control hinges on an asymmetry of knowledge between authority and those subject to it, and, as we've established, in the context of visuality this asymmetry takes shape around depictions of who is in the position of *producing* meaning and those who are in turn produced *as* meaning. This is cultivated through various means of appropriation, concealment and capture that are unique to the means of representation dominant within a given complex, each of which produces and deploys its own "apparatus to hide the apparatus."[150] For the counterinsurgent complex, the dominant apparatus of concealment is the user interface.

Typified by the internet browser (but here expanded to include standalone applications, operating systems and platforms) the interface is in essence a protocological "machine that uses a set of instructions to include, exclude, and organize" data and aestheticize it in such a way as to conceal the operations behind its visualizations.[151] Interfaces thus transform the user experience of networks from the otherwise "unnerving experience of radical dislocation—passing from one server in one city to a server in another city" into the comparatively pleasurable sensation of uninterrupted movement aestheticized under the colloquial moniker of *surfing*.[152] Galloway of course reminds us that the experience of seamlessness and continuity within the interface is in reality an illusion created through strict adherence to an array of technical standards designed to

---

reauthorization, writing that "[as long as we] see ourselves as defeated victims, we can refrain from having to admit that we are short on ideas—or that the ones we have seem unpopular, outmoded. Thus, we *need* a strong, united enemy." Jodi Dean, *Democracy and Other Neoliberal Fantasies: Communicative Capitalism and Left Politics*, 1st unedited edition (Duke University Press, 2009): 6.

make the network as intuitive as possible, reducing barriers to action while rendering invisible the physical and protocological apparatuses which make that action possible. As such, the interface is emblematic of Visuality's denial of the right to look, insofar as it is the primary mechanism for concealing both data and the means by which networks, platforms, and algorithms visualize *interested* realities from it.

Because the current economic model of the internet revolves around platforms, they together form what Ulises Mejias and Nick Couldry have called a new "social quantification sector" where the breadth of human activity is seamlessly converted into data to be directed toward the complementary goals of value generation and control.[153] Once extracted, this data is often sequestered via an array of legal and technical closures, leading to the expropriation of social knowledge/data as private property and the proliferation of technical black boxes in the form of the opaque machine learning algorithms and user interfaces that serve to further obfuscate the production of meaning behind the illusion of a seamless input/output relation. The right to look is thus denied, with platforms claiming the authority to algorithmically and inconspicuously arrange "the relations of the visible and the sayable,"[154] despite (and, paradoxically, as a result of) mass participation.

This arrangement echoes the foundational logic of what Mirzoeff terms Visuality 1, as that mode of visualization depending on "a servile class, whether formally chattel slaves or not, whose task it is to do the work that is to be done and nothing else."[155] Visuality thus renders this sub-class operational not only through our labour, but in our being denied access to the means of visualization made possible by and through our labouring. Even as we perform the work of content production, communication, and the countless other social performances which take place on and through platform services, the authority to interpret, render meaningful, and govern

those actions is both centralized and concealed: “for us, there is nothing to be seen.”[156] As a primary vector of social, cultural and economic activity, platforms have transformed the performances of everyday communication, knowledge production, and exchange into the raw material necessary for authority’s visualization. Through the capture of life as data and its encapsulation within the black boxes of technical and legal obfuscation, counterinsurgent visuality is thus able to legitimize the totality of its authority by creating a condition wherein “it alone can visualize the divergent cultural forces at work in a given area and devise a strategy to coordinate them.”[158]

The technological and media conditions native to counterinsurgent visuality present a novel hurdle to projects of counter- (or insurgent) visualization. First, the means of visualization have become increasingly *in*visual, rooted in the operationalization of large-scale assemblages of data that “cannot be seen by an observing ‘subject’ but rather [are] enacted via observation events distributed throughout and across devices, hardware, human agents and artificial networked architectures such as deep learning networks.”[159] Comprised of non-representational data, the visualized field of counterinsurgency creates a “3-D rendition of the insurgency that corresponds to the counterinsurgent’s experience of space in a grid accessible only to the ‘commander,’”[160] and invisible to the unaided eye. This “3-D rendition” can be likened to the invisual operations of a generative model’s latent space. As discussed in the literature review, latent space is not a neutral abstraction but a statistical regime of visibility/legitimacy, where proximity within a vector field encodes what is knowable and retrievable. Like the counterinsurgent map rendered for the commander’s eyes only, a models’ latent space governs the field of possibility from a distance—producing meaning and action through logics that themselves remain unseen. This discorrelation[161] of visuality from the perceptual and imaginary

faculties of those who are ultimately subject to authority has led to a situation where power is produced through distributed events and processes made increasingly unavailable for countervisualization through combinations of both technical and legal prohibition and the cultivation of "invisibility in support of a digitized surveillance and command structure."[162]

Second, counterinsurgency differs from previous complexes of visuality in that it either produces or co-opts its own counters as a necessary condition of its authority: "Whereas Carlyle offered the Hero and his visualization as the only defense against chaos, the counterinsurgent *requires* chaos, or at least its possibility, as the means of authorization in all senses."[163] This is particularly visible in the functioning of platform algorithms designed to increase data extraction and ad revenue by modulating content visibility in ways which specifically target and farm the attention of individual users. One result of this unwavering commitment to engagement is that the curated representations offered by platform algorithms can often lead to the active shaping of politically oppositional subjectivities as a direct result of their design. To this point, Jason Buel has argued that, by optimizing for engagement "at any price," platforms facilitate a particular kind of subject formation in which algorithmic control is less concerned with the meaning of specific representations than with modulating their circulation—ensuring that even oppositional content is absorbed into viewing habits in ways that keep the prospect of meaningful political transformation below the threshold of possibility.[164] In this way, at least in theory, algorithms are able to perform and sustain counterinsurgency by capturing insurgent visualization as value, while at the same time modulating their visibility in ways that "automates the [visualization] of a capitalist realism that fundamentally limits the potential of social movement discourses to bring about change."[165] As Mirzoeff put it, the presently dominant model of culture "might be described as that process by which the excess of visuality in Visuality 2 is made available as part

of the modernizing process to Visuality 1."[166] This mirrors Mark Fisher's account of capitalist realism as a mode of cultural production in which subversion is no longer simply incorporated after the fact, but pre-emptively formatted, redirected, and/or neutralized. Rather than responding to insurgency through repression or assimilation, platforms enact a form of what Fisher called 'precorporation'—the "pre-emptive formatting and shaping of desires, aspirations and hopes by capitalist culture."[167]

The objective of these algorithmic techniques is therefore not to assimilate the insurgent, or to ameliorate the conditions which lead to insurgency, but to manage insurgency as the precondition for their regimes of classification, separation and control, all the while preserving an underlying political economic hegemony. This is perhaps most transparent in the phenomena of algorithmic apophenia or "inceptionism," where the drive of algorithmic systems to identify and manage patterns reveals itself to be so strong as to hallucinate them in their absence.[168] Mirzoeff marks this indifference "to what is known or unknown" as a source of counterinsurgent visuality's resilience, noting that the success of the intensified complex is measured solely by the permanent continuation and re-authorization of its apparatus of control.[169] The primary discursive activity of counterinsurgent visuality thus shifts from legitimizing its regime of separation, to the conscription of countervisualization in the technological regulation of its own visibility, informing and effecting modulations intended to keep insurgency either out-of-sight or contained so that neoliberal governance can proceed unperturbed.

The Crisis of Visuality

Writing on its shortcomings as military strategy, Mirzoeff invokes the United States' multiple failed exercises in "nation-building" as emblematic of a developing crisis of Visuality 1, where the social and political realities it seeks to impose have become increasingly incoherent. Despite the appearances of partial success—as in Iraq, where the Bush administration's campaign briefly presented the illusion of stability—counterinsurgency efforts have consistently failed to deliver basic conditions of legitimacy. These failures, Mirzoeff argues, are symptomatic of a deeper, more qualitative issue. In the era of globalization, transnational migration, and digital communications, modern counterinsurgency increasingly lacks the political coherence necessary to define itself "beyond the classification and separation of the insurgent"—a fundamental distinction which has itself become increasingly difficult to resolve.[170] Counterinsurgency, having become so "enmired in 'security' as to have lost a sense of purpose,"[172] now consistently fails to hold even the imperial fantasy of the "host population" as a cohesive and "healthy" social body structured around the worn (but still very present) unifying images of the market and nation. Increasingly driven by data collection and statistical modelling, counterinsurgency shapes an aesthetics of prediction over interpretation, maintaining discursive control while keeping the means and perspective of its visualization undecided or obscured. So invested in the naturalization of asymmetry, it produces a culture requiring its permanent intervention, trading the promise of progress, stability, or renewal for a perpetual re-authorization of counterinsurgency as the management style best-suited to chaos.

In what Mirzoeff speculates could be its "final intensification," counterinsurgent visuality proceeds to "render the visible invisible, even within the zone of those authorized to see."[173] It achieves this, he says, through the cultivation of chaos through overexposure and ambiguity so that 'there is nothing to be seen,' not because it is concealed, but because "it cannot be decided

what has been seen." What results is a set of conditions where authority's legitimacy is derived not from mastery, but the "ability to ignore the constant swirl of imagery and persist with a "vision" above and beyond mere data."[174]

But Mirzoeff suggests that visuality *itself* becomes "visible" at its present point of intensification. The conditions of chaos and invisibility* surrounding the creation and circulation of representations online have led to a situation where it is increasingly difficult to contain visualization "within the circle marked out by the police,"[176] and authority's *un*reality is made readily apparent. What emerges is the picture of a complex of control no longer tethered to a stable political project, and thus stripped of its legitimating claims. While counterinsurgency has been marked by a general indifference to what is known, the algorithmically-driven proliferation and aggregation of countervisualization for profit has in recent years bled into political life, seeding growing distrust and disillusionment with the apparently vacuous underpinnings of neoliberal governmentality.

Conclusion

Part One of the dissertation has sought to establish the context necessary for situating generative AI in relation to broader genealogies of visuality, media, and power in order to later present its emergence and subsequent capture as an intensification of counterinsurgent control.

* Paradoxically, the invisibility represented in the analogy of the "black box" is concurrent with the unprecedented possibility of *seeing* Visuality itself, as its power has never been more conspicuous than at the point of its translation into code.

This argument unfolds against a backdrop of mounting instability in the dominant visual order, as long-standing techniques of legitimization struggle to contain the excesses they once managed.

Stoked by platform architectures which sought to aggregate and appropriate their sentiments as value, insurgent discontents have been seized upon by political operatives as raw material for countervisualization, leveraging the algorithmic underpinnings of the platform to metastasize 'previously-controlled' *ressentiment* into a world historical force. In the face of this crisis, there appears to be a general lack of will and ability for authority to re-imagine a coherent and compelling alternative to the further entrenchment of discursive regimes that are either neoliberal, fascist, or—as is more often the case—both. In this absence, we are instead witness to further intensifications of counterinsurgency, exemplified in the domain of information and communication by the rise of disinformation studies following the success of the Trump and Brexit campaigns, the various regulatory, technical, and epistemic remediations that field continues to produce, and its current instrumentalization in struggles that increasingly turn inward—redirecting the counterinsurgent logic against new domestic adversaries, including dissenters and critics aligned with the very liberal-democratic norms these interventions were originally intended to defend.

Here and in the literature review, I have argued that machine learning models—especially generative ones—do not simply reflect reality, but actively construct and contain it, and as such embody a counterinsurgent logic attuned to the modulation of visibility and legitimacy under conditions of perceived disorder. This renders them especially potent in the current moment, where challenges to institutional authority have triggered renewed efforts to manage insurgency through technological means. It is in this context of crisis that Part Two begins. Turning from

visuality's history to its contemporary dislocation, the next section explores how the rise of disinformation discourse functions as a mechanism for re-legitimizing authority and control. There, I examine how generative AI is caught in a dual bind: positioned as both a threat and a tool for restoring order, and how its securitization folds it into broader projects of epistemic stabilization and counterinsurgent governance.

PART TWO

Crisis: Aggregation and the Securitization of Generative AI

Introduction

Where Part One traced a genealogy of visuality as a historically-adaptive ordering of the social, Part Two turns to the multivalent crisis—epistemic, discursive, and political—which makes the intensification of counterinsurgent visuality not only possible, but necessary from the standpoint of establishment authority. This section argues that the crisis of visuality, identified by Mirzoeff in 2011, finds its contemporary expression in the overlapping crises of neoliberal legitimacy and digital governance, where the increasing visibility and aggregation of insurgency disrupts homeostatic regulation and compels intensified control over content, communication, and the generative models that increasingly mediate them.

Against narratives framing crisis and disorder as the products of disinformation, conspiracy, or foreign meddling, Part Two presents the argument that what is at stake is not the collapse of truth or objectivity, but a loss of establishment control over the terms of legibility, as new communication technologies unsettle visuality's ability to aestheticize legitimacy. This is not a crisis of information quality, as dominant narratives suggest, but of *aggregation*: the increased capacity of publics to cohere, however unevenly, into politically legible formations outside the protocols of visuality. Central to this argument is Wendy Chun's theorization of homophily—the tendency of like to associate with like—not as a natural social fact, but as an operational principle embedded in the platform algorithm.[1] Where once disinformation

campaigns and technocratic management could defer dissent, "homophilic clustering technologies" have rendered dissent newly visible. Rather than merely producing echo chambers, I contend that this clustering enables the formation of *insurgent aggregates*—political subjects that emerge from shared disaffection and circumvent established mechanisms of legitimation. In this view, platform culpability for the crisis does not extend from their role as vectors of disinformation, but in their unintended capacity to facilitate aggregation in ways that disrupt the homeostatic equilibrium of counterinsurgent visualization.

In this context, popular disaffections are pathologized as disinformation, and the infrastructures of communication become the site of *securitizing moves* that reframe both dissent and the emerging technologies of generative AI as existential threat. To this, the prevailing establishment response has been to intensify counterinsurgency—no longer able to relegitimate discursively, the order turns instead to techniques of regulatory, technical, and epistemic containment. As generative AI is positioned as both a threat and solution, it is increasingly shaped to serve as a tool of *disaggregation*. This section proceeds in three movements—crisis, aggregation, and securitization—to trace the conjuncture that has allowed "safe" generative models to emerge as instruments of intensified counterinsurgency. By mapping this crisis discourse to its political functions, Part Two builds the bridge between the groundwork laid in Part One and the technical and environmental analysis of the "safe" generative model in Part Three, situating crisis as the terrain on which counterinsurgency intensifies.

Crisis

In my understanding, the crisis of visuality described by Mirzoeff points to an inevitable faltering of counterinsurgent visualization, marked by the complex's increasing inability to aestheticize a cogent social and political reality in legitimation of its classificatory and separative regime. While this has of course been reflected in the many strategic and political failures of COIN-based western foreign policy since the Vietnam war, a similar crisis has likewise questioned its effectiveness as the complex discursively sustaining the neoliberal establishment. While accounts of this failure vary, the 2008 financial crisis is now generally understood as the moment when the already tenuous sense of legitimacy around neoliberalism, which had for years "provided ideological cover"[2] for rampant privatization and the unfettered accumulation of wealth, collapsed.[3] While one would expect that the spectacular undressing of an ideology which dominated political economy for the better part of fifty years might trigger a period of introspection, the core tenets of neoliberalism have instead continued to inform Western policy in the years following. Absent the pretense to legitimacy afforded by the rapid economic growth and stability of the eighties, nineties and early aughts, the enactment of policies based on deregulation and the privatization of public institutions persists, as Mark Fisher put it, "no longer as part of an ideological project that has a confident forward momentum, but as inertial, undead defaults."[4]

Fisher famously described the deprivation of social and political imagination necessary to allow for neoliberalism's continued dominance as being symptomatic of a now pervasive "atmosphere" of its *inevitability*. This atmosphere, which he called capitalist realism, emerges as an effect of the neoliberal 'business ontology' in which "it is *simply obvious* that everything in society, including healthcare and education, should be run as a business."[5] According to this reading, neoliberalism's seeming inexorability, coupled with the sense that "it is now impossible

even to imagine a coherent alternative to it,"[6] has sparked a global trend toward politics of resignation formed around the post-political[7] mantra of 'there-is-no-alternative' (or TINA), leading to a condition wherein political discourse itself shifts from the articulation and contestation of ideas to the framing and defense of neoliberal policy as technical and therefore beyond debate. As they so often do, the Germans even coined a singular term to encapsulate the new, muddy feeling: *alternativlosigkeit*—alternativelessness. I contend that capitalist realism and the post-political technocratics of TINA together index the evolving crisis of visuality described in Mirzoeff, as a situation where the dominant order has become either incapable or disinterested in justifying its visualizations as a rational discourse through politics, engaging instead in the technocratic management of an increasingly unpopular and dysfunctional complex. Unable to reassert its realities as fact, the complex of visuality thus fails to naturalize its operations and, in the absence of alternatives, intensifies counterinsurgency in a bid to sustain the formation despite growing challenges to its legitimacy. As the visibility of dissenting information across the digital public sphere has increased over the last decade, its classification, separation, and suppression under the umbrella of "disinformation" has emerged as a central feature of this intensification.

Contrary to conventional understandings of the 'disinformation order'—which generally attribute current epistemic and political destabilization to the effects of social media, foreign interference, and the cognitive shortcomings of gullible citizens[8]—scholars in the growing field of critical disinformation studies locate the phenomena's roots elsewhere. In their now foundational paper, "A Brief History of the Disinformation Age: Information Wars and the Decline of Institutional Authority," W. Lance Bennett and Steven Livingston complicate mainstream accounts by linking the current state of 'information disorder' to a crisis of institutional legitimacy induced through decades of official disinformation (read: *public*

relations) propagated in support of free-market economics and limited government. Bennett and Livingston's account argues that the crisis is not a spontaneous effect of digital media but rather the product of long-term structural transformations in the public sphere, particularly the capture of core institutions by corporate interests and political elites who have increasingly come to rely on disinformation to advance otherwise unpopular policy agendas. Characterized by partisan think tanks, corporate spokespeople, government officials, and political action committees subjecting information to varying degrees of "spin," this communication environment is held up as having gradually undermined the credibility of many mainstream political and media institutions.[9]

According to this explanation, despite the widespread capture of politicians, governing institutions, and mainstream media outlets in service of rationalizing neoliberal deregulation, public disaffection with each has grown over the last fifty years as the disparities between political rhetoric and actual policy outcomes have become increasingly difficult to reconcile. In this view, declining trust in gatekeeper institutions can be traced to a realization among the public that many have in effect been "damaged by information warfare intended to limit [their ability] to regulate their own social and economic affairs."[11] Absent media and political representation which credibly reflects their interests, a growing number of citizens, disillusioned with mainstream political discourse, thus begin seeking alternative information and narratives which speak to the emotional realities of their increasing political and economic marginalization.

Aggregation

Prior to the 2008 financial crisis, the neoliberal project was able to effectively suppress opposition to the technocratic management of political and economic affairs with the help of an information ecosystem formed around and shaped by a handful of authoritative institutions, which together served as the de facto gatekeepers of political discourse.[12] Following the crisis, however, discontent with the political-economic status quo would come to a fever pitch, eventually leading to the emergence of various 'political monsters' which Bennett and Livingston describe as the "unintended outcome" of the political and discursive work that created the regulatory conditions which fomented the collapse.[13] In the United States, specifically, anti-establishment sentiments boiled over into a deeply fractured political environment, with the Occupy and Tea Party movements emerging to represent the shared discontent of citizens from opposite sides of the political spectrum. In the years since, the situation has only deepened as new (and arguably more structurally disruptive) threats have emerged in the form of right-wing populist movements characterized by anti-globalist politics and "interestingly selective attacks on elite economic rule."[15]

After 2008, the proliferation of anti-establishment protest and political organization was further stimulated by the (then relatively new) technology of social media, a paradigmatic shift in communications which fatally disrupted previously-dominant systems of discourse management. So follows a core tenet of critical disinformation studies: the decentralization of the information ecosystem has not *produced* our current state of political and epistemic discord (a claim central to mainstream disinformation rhetoric), but simply restructured the public sphere in ways that reallocate attention and visibility to an assortment of previously-suppressed political sentiments. By virtually eliminating barriers to participation, new social media platforms revealed the attitudes of publics previously unrepresented by established political organizations, and, through

processes of algorithmic sorting designed to increase engagement and advertising effectiveness, inadvertently "provided opportunities for people holding these opinions to find each other, distribute information, […] and organize their challenge to the status quo."[17] In a 2021 article for Harper's introducing critical disinformation to a wider audience, Joseph Bernstein poses the important (perhaps rhetorical) question of whether social media has created "new types of people" or simply revealed "long-obscured types of people to a segment of the public unaccustomed to seeing them?"[18] To offer an answer in the parlance of this dissertation: as a side effect of decentralization and the algorithmic governance of communication flows, social media platforms have (as an unintended side-effect) allowed for the unchecked *aggregation* of anti-establishment insurgency, gathering previously atomized groups of people around shared disaffection with the social, political, and discursive formations of Visuality 1.

Analyzing the politics of digital networks in her book *Discriminating Data: Correlation, Neighbourhoods, and the New Politics of Recognition*, Wendy Chun uses the concept of homophily (the principle that similarity breeds connection) to describe how platform algorithms deliberately "create agitated clusters [or aggregates] of individuals whose angry similarity and overwhelming attraction to their common object of hatred both repel them from one another and glue them together."[19] According to Chun, the recommender algorithms which breed connection on social networks act as "homophilic clustering technologies," engineering online communities held together by a combination of shared interest and/or antagonism for the cultural, political, or racial other.[20] Having first *disaggregated* the (supposedly) homogenized national cultures produced and sustained by mass media, algorithmic filtering then reorganizes the public sphere into the innumerable clusters of like-minded interaction and ideological-affirmation described by the now familiar "filter bubble" and "information silo."[21] Following this, the complementary

effects of polarization and aggregation can be read as primary goals of the platform's regime of governance, which sought to captivate and restructure its mass of users into demographic 'packages' to be sold to advertisers.

Despite the professed aspirations of tech leaders to craft algorithms which "improve our well-being and happiness" and support "meaningful interactions" between friends and family,[22] the homophilic aggregates created by platforms are as often shaped around kernels of upset and division,[23] and therefore easily operationalized for political or economic gain. Chun elaborates on this point, noting that homophilic clustering techniques do not end with the formation of discrete groups, but enable them to then be strung together like an affectively-charged "strand of pearls," ripe for political mobilization.[24] While common examples of this aggregation of isolated users into new collective subjects often include the resurgence of right-wing extremism, Covid-19 denialism and the (related) QAnon phenomenon, they share an origin as politicized expressions of widespread disillusionment with institutional authority and a state of political and economic affairs that feels increasingly untethered from the concerns of everyday people. As Chun points out (via Sara Ahmed), the politics of binding isolated "I's" into a threatened "we" (and thus susceptible to control) requires the identification/creation of an Other, through which *hatred* (of the other) can be laundered into *love* (of one's own)—as homophily.[26]

Up to the point of crisis, the neoliberal project managed to effectively operationalize homophily through disinformation in support of its policy objectives, focusing the ire of disaffected publics onto carefully-crafted racial and class others who, so it went, disproportionately benefited from unearned and superfluous government support. In more recent times, however, the tactic of authorizing neoliberal policy via the "divining and amplifying [of] perceived stigmas, and [then] consolidating them together around a common 'enemy'"[27] has

begun to backfire. As Bennett and Livingston have observed, this strategy of "fanning the flames of hatred and division" has proven volatile, and as changing media conditions have made this work increasingly visible, ultimately reveals the fundamental contradiction of neoliberal governance: most of the public "would not buy it on its own terms."[28] As such, the continued use of 'official disinformation' to sustain elite consensus has instead had the effect of creating new and formidable opposition movements which are not easily controlled via existing strategies of political containment.

Having lost the ability to redirect the corrosive energies produced by decades of political and economic disenfranchisement, the forces of neoliberalism have themselves increasingly been countervisualized on social media as the parasitic other threatening the interests of the people. As homophilic clustering technologies, platform algorithms have thus functioned as intended: identifying and supporting 'meaningful connections' between large numbers of politically-alienated users, in turn enabling the relatively unchecked aggregation of anti-establishment insurgency online. Of course, the most talked about manifestation of this effect in democratic societies has been the resurgence of right-wing populists who exploit homophily to string clusters of agitated citizens into a bloc of political and electoral consequence, but opposition has emerged from across the social and political spectrum, and the growing number of movements employing social media as their chief means of outreach and organization have included the left-leaning #BlackLivesMatter, #Occupy and #FridaysForFuture as much as the more anti-liberal and xenophobic #Leave or #MAGA campaigns. Despite their drastic differences, I suspect what ties many contemporary protest movements together is a shared disillusionment and distrust in the institutions and principles of a post-political establishment depriving them of agency, thus allowing deregulation and the privatization of public assets and services to proceed

unencumbered by politics. As such, insurgent aggregation and the related phenomena of mis/disinformation must be understood less as issues of information quality than as expressions of deeper social and political tensions, and therefore not properly addressed through technocratic means, an approach that Jungherr and Schroeder have described as "the doctor only treating a patient's symptoms while missing the cause of the disease."[30]

By characterizing disinformation as an effect of media decentralization and the attendant degradation of information, mainstream discourses prioritize technical solutions to problems that are fundamentally social and/or political in nature. Such an approach is, of course, characteristic of the capitalist realist politics of no alternative where, rather than serving as the "pretext for the emergence of different ways of living,"[32] narratives of political and epistemic crisis serve to authorize newly intensified controls. Absent the possibility of redress, increasing tensions between the neoliberal movement and its 'political monsters' therefore require "more creative management" of the democratic process, shaping flows of (dis)information so as to capture and/or *disaggregate* opposition and "further guard against any of these movements or parties threatening business interests."[33] Quoting former Israeli Foreign Minister Moshe Dayan on the issue of Palestinian statehood, Mirzoeff claims that counterinsurgency as a mode of governance substitutes the political question 'What is the solution?' with the technocratic 'How do we live *without* a solution?'.[34] So far, the answer—which seemingly applies to the digital public sphere as much as the Palestinian territories—has been to intensify the complex which sustains authority. In the context of our discussion on generative AI, this involves the securitization of new information and communication technologies through their discursive construction as existential risk.

Securitization

Originating in international relations, securitization describes the process by which state and institutional actors transform issues of *politics* into matters of *security* in order to authorize extraordinary intensifications of control. I believe this concept is critical to unpacking how rhetorics of disinformation and epistemic collapse contribute to the construction of ICT's (generative AI, particularly) as threats to the security and existence of democratic societies. As a frame of analysis, securitization allows us to scrutinize the discourses which function to move issues from the realm of politics into the realm of security, thus presenting them as being beyond deliberation. According to the model introduced by Buzan, Wæver, and de Wilde, this process marks an issue's movement along a spectrum from either 'nonpoliticized' (excluded from public or political debate) or 'politicized' (subject to deliberation and requiring governance) to 'securitized'—meaning it is framed as an *existential threat* to the society, and thus "requiring emergency measures and justifying actions outside the normal bounds of political procedure."[35]

Through adopting a definition which emphasizes a specific rhetorical structure over direct ties to 'security' as it's traditionally understood, analysis may also be expanded to issues and phenomena said to constitute threats beyond the restrictive confines of the military-political to include particular patterns of communication and knowledge production.[36] In order to distinguish acts of securitization from regular political discourse, analysts typically look for rhetorical structures and devices unique to the process, specifically the evocation of existential threat and related utterances as a means to produce an atmosphere of urgency. By staging a given issue as existential (to a group, nation, system, species, etc.), security actors attempt to create the sense that if the problem is not dealt with swiftly and decisively, it will weaken the host society

to such an extent that future remedy will be either ineffective or impossible, thereby authorizing extraordinary measures to bring it under control. Finally, the securitization process is further distinguished from regular political discourse by the successful introduction of measures which almost invariably circumvent politics, allowing for the enactment of new policies and controls which shirk oversight, unilaterally reallocate public resources, and/or infringe on otherwise inviolable rights.[37]

In light of this definition, it is clear to see the process of securitization at work in the popular and institutional discourses surrounding generative AI, wherein the presentation as threat is made explicit through their regular invocation in relation to a state of epistemological and democratic crisis with grave consequences for the future of democratic society. Following the political upheavals of 2015 and a rapid development of generative techniques culminating in the public release of ChatGPT in 2022, these rhetorics have ramped up and spread widely through a near-constant stream of newspaper op-eds, think-tank reports and international conferences on AI, disinformation, and their combined contributions to a state of affairs that some commentators have popularly called the "infocalypse."[38] Here is a small, but representative sampling: a 2022 report from the Center for Security and Emerging Technology at Georgetown University, unambiguously titled "The Existential Threat of AI-Enhanced Disinformation Operations," which warns of an "inevitable" (but as yet unrealized) future where new AI capabilities "erode trust in democratic governance," sow doubt in "the possibility of truth in public life," and lead societies down a path toward "mob majoritarianism";[39] a 2024 opinion piece authored by the editorial board of the Globe and Mail, urging legislators and their constituents to support the speedy criminalization of "malicious deepfakes" (an ill-defined combination of terms) as generative AI threaten to turn social media "into so great a threat to the public good that it could

make the current era of online disinformation seem like a quaint prequel to the apocalypse";[40] a round-table discussion at the 2023 meeting of the World Economic Forum in Davos titled "The Clear and Present Danger of Disinformation," where speakers from government, media and academia repeatedly cite AI as a central contributing factor to increasing political and epistemic instability;[41] and US Vice President Kamala Harris who, at the 2023 Bletchley Summit on AI Safety, punctuated her wide-ranging speech with the rhetorical question, "when people around the world cannot discern fact from fiction because of a flood of AI enabled myths and disinformation. I ask, is that not existential for democracy?"[42]

Importantly, the securitization framework notes that a discourse which declares an existential threat (the *securitizing move*) does not, on its own, satisfy the criteria of securitization. In order for an issue or object to be effectively securitized, the securitizing move must successfully compel its audience to accept the threat as such and, in turn, the imposition of extraordinary measures. As Buzan, Wæver, and de Wilde put it, successful securitization requires "the intersubjective establishment of an existential threat with a saliency sufficient to have substantial political effects"—that is, it must gain enough resonance with its audience to justify measures that "would not have been possible" without the securitizing move.[43] Accordingly, in order to argue that generative AI has been (or is in process of being) securitized, one must provide evidence that the audience (whether that be the public, policymakers, industry, etc.) has accepted the securitizing actor's premise such that they might tolerate the violation of established rules and procedures in the course of addressing the proposed threat.

Acknowledging that much of the available evidence comes from data gathered by organizations which might be classified as securitizing actors themselves,[45] surveys reveal high (and increasing) levels of concern about generative AI among the public and other stakeholders.

Based on a survey of "1,490 experts across academia, business, government, the international community and civil society," the World Economic Forum's 2024 *Global Risks Report* identifies "AI-generated misinformation and disinformation" as the most severe risk* to global security over a 2-year period (ahead of climate change, armed conflict, and disruptions to critical infrastructure), coming second only to climate change related risks when the timeline is extended to 10-years.[46] Similarly, a 2024 study of trends in Canadian's internet use revealed that more than half of those surveyed were worried about the epistemic effects of generative AI, listing the "spread of fake images or videos," "spread of mis/disinformation," "Insufficient regulations/controls on its use" and "potential to disrupt human society" as chief concerns.[47]

Following this, additional data generally reflects a steady rise in support for interventions that would have been hard political sells before securitization. Perhaps most illustrative of the point are a series of survey's performed by the Pew Research Center indexing rising support for the imposition of measures which "restrict false information online, even if it limits people from freely publishing or accessing information."[48] Comparing data from polls conducted in 2018, 2021, and 2023, the studies show support for government intervention increasing from 39%, to 48%, to 55%, with even greater support for intervention by tech companies at 56%, 59%, and 65%, respectively.[49] According to the research, this rise in support for restrictive measures also parallels a steady decline in respondents who believe that "freedom of information should be protected even if it means false information can be published"—42% in 2023, down from 50% (2021) and 58% (2018).[50] Buoyed by the increase in political support, recent years have seen a

* Tellingly, the WEF report—which classifies disinformation as a 'technological' rather than social or political problem—comes at a time when the relevance of the organization is itself being unsettled by changing political and economic tides. Caitlín Doherty, "At the Summit," *Harper's Magazine*, accessed April 19, 2025, https://harpers.org/archive/2025/02/at-the-summit-world-economic-forum-davos-caitlin-doherty/.

proliferation of institutes, commissions, and ‘task forces’ devoted to addressing the cause and effects of ‘information disorder’ through research, policy recommendations, and tailored technical interventions.

A final characteristic of securitization is that the discursive construction of threat is generally grounded not by any objective measure of security, but instead emerges as the effect of an *intersubjective* process whereby the security actor influences its audience to accept its understanding.[51] Put another way: “securitization is a rule-governed practice, the success of which *does not necessarily depend on the existence of a real threat*, but on the discursive ability to effectively endow a development with such a specific complexion.”[52] In order for an issue or object to count as securitized in the way we’re applying it here, the empirical evidence in support of its threat is not only secondary to the securitizing move, but is often slim, shaky, or non-existent. For example, climate change, while it satisfies the other conditions, would not qualify as *securitized* in this sense, as the objective reality of its threat is borne out by overwhelming empirical evidence—barring subjective disagreements about what conclusions may be drawn from that data. On the other hand, existing research on the apocalyptic epistemic consequences of generative AI suggest only modest effects. Mainly, research shows that generative AI do not meaningfully exacerbate problems of mis- and disinformation beyond what was possible with prior technologies,[53] nor do they—on their own—increase the demand for misleading or ‘harmful’ information.[54] The centrality of largely unsupported alarmist rhetorics to the securitization of generative AI and other ICT’s is further evidenced by studies correlating the over-emphasis of risk with increases in audience perceptions of danger and, in turn, increased support[55] for “heavily restrictive regulation of speech in digital communication environments.”[56]

In light of these conditions, some academics and commentators have proposed that we might constructively think of the recent concern over disinformation as a new *moral panic*, with generative AI standing in as the 'folk devil' that threatens social order by initiating and/or accelerating structural transformations of the public arena.[57] As elaborated in the literature review, the moral panic framework involves the formulation, exaggeration, and communication of threats attributed to the folk devil by an alliance of 'panic actors' or 'primary definers,'[60] resulting in the arousal of public concern and the eventual introduction of new control measures. Jack Bratich updates this framework for the present, suggesting that such panics are no longer simply techniques of stabilization but have themselves becomes "strategies in a combative context," deployed by threatened or collapsing regimes to reassert control in the absence of legitimacy.[61] Following this, the moral panic surrounding disinformation may be understood not simply as a response to epistemic instability, but as a technique of relegitimation, and a prelude to the securitization of generative AI as an emergent technology of communication and knowledge production. With generative AI successfully securitized, an assortment of extraordinary regulatory, technical, and epistemic control measures—developed under the pretense of crisis—intensify counterinsurgency as the dominant mode of visualization, shaping the new technology as a powerful means of modulating and/or neutralizing insurgent patterns of knowledge, communication, and political action that threaten to destabilize the existing political economic order.

Conclusion (Counterinsurgency)

In the context of crisis, the rhetorics of securitization and moral panic have increasingly employed the language of war as a primary interpretive frame for understanding information and politics, often to the effect of conflating political dissent and alternative epistemological formations with the interference of foreign and malign domestic actors under the banner of the 'war on disinformation.' Characterizing it as the discursive projection of a crisis policing "now *martialized*," Jack Bratich has suggested that the current moral panic thus reflects the aspirations of a beleaguered neoliberal establishment now engaged in a "war of restoration," the express goal of which is the "[re]unification of society around a political spectrum dominated by a center."[64] Responding to structural changes which had effectively eliminated barriers to access and participation in the public arena, the concerns about disinformation and the potentially exacerbating effects of artificial intelligence might therefore be seen as reflecting a willingness on the part of traditional and institutional gatekeepers to "attribute any visible deviance from the status quo with the potential to damage democracy."[65] Such a tendency threatens the imposition of broad countermeasures which effectively weaponize infrastructures of communication and knowledge production against political and epistemological discord *as such*, restoring legitimacy and 'order' via the cultivation of a contrived unity, definable only in contrast to *divisiveness* as a quality of political and epistemological contestation.

Whereas previous moral panics may have served to restore a 'threatened' social order through its triumph against--and ultimate reintegration of—the folk devil, Bratich reminds us that the current conjuncture is unique. "Rather than being able to hegemonically relegitimate," he writes, the moral panic over disinformation has instead produced "a set of actions drawn from counterinsurgency warfare that *return politics to war*."[66] Having passed "a tipping point of irrecuperability,"[67] the order thus seeks recourse through more repressive means—not only

censorship, but entirely new techniques for the neutralization and management of domestic dissent, exercised through a dispersed network of tech companies, intelligence agencies, media outlets and establishment politicians.

Defined by David Kilcullen, a leading military theorist, as "specific measures, tailored to the environment, to suppress a particular insurgency and strengthen the resilience of a particular threatened society and government,"[71] counterinsurgency—on its face—seems a natural response to the disorder represented in the figure of the disinformation 'devil.' However, being a primarily military-strategic (rather than political-economic) doctrine, its explicit embrace reflects the persistent tendency to visualize the social as a site of management and control over and above the more sticky undertakings of social and political (re)legitimation. According to modern COIN doctrine, insurgencies are not defined by the *presence* of preexisting elements like declining trust, disillusionment, or dissenting citizens, but their organization into *new patterns* which run contrary to the interests and stability of an incumbent regime. Kilcullen explains that, "because the elements of insurgency are preexisting but the *pattern of interaction* is new," the counterinsurgents' goal is not to eliminate the individual elements or nodes of an insurgency, but to rearrange their patterns of interaction into a "stable and peaceful 'system state'."[72] In the terms of our argument, this 'system state' takes the form of what Bratich calls "a unity without content,"[74] here understood as the cultivation of 'stability' through the neutralization of insurgent patterns while leaving the underlying elements unaddressed. Under this model, preexisting elements which risk aggregating as insurgency are largely suppressed or funneled into more easily manageable directions, a tendency perhaps reflected in what Anton Jäger has more recently referred to as *hyper-politics*—a state of affairs characterized by fervent popular

political action which, occupied with the atomized policing of identity and interpersonal mores, fails to catalyze meaningful opposition to the defaults of neoliberal governance.[75]

Having discursively reshaped the terrain of information as the site of war, the moral panic around disinformation now securitizes and conscripts platforms, artificial intelligence, and other technologies of communication and knowledge production into the intensification of counterinsurgent visualization and control. In the following section I will argue that—shaped by a complex of extraordinary regulatory, technical, and epistemic countermeasures—generative AI is currently developing as a powerful agent of this intensification, as the intentional deployment of the model's inherently counterinsurgent operations against political and epistemic insurgency. Under the auspices of restoring order and integrity, the resulting "safe" generative models threaten to naturalize unprecedented levels of control, with significant implications for the expression, communication, and organization of political alternatives.

PART THREE

Intensification: Disaggregation and the 'Safe' Generative Model

Introduction

If Part One developed a framework for understanding media technologies as material supports for co-constituted complexes of visuality and power, and Part Two situated the overlapping crises of visuality and neoliberal legitimacy as the terrain for counterinsurgent intensification, Part Three turns directly to the object at the centre of this dissertation's intervention: the "safe" generative model. Building on the earlier claim that connectionist AI systems are active in the operationalization, intensification, and naturalization of counterinsurgent visualization, this section unpacks the positioning, development, and deployment of the "safe" generative model in relation to the broader political project of establishment relegitimation.

As earlier sections have shown, counterinsurgency differs from prior regimes of visualization in that it imagines the social not as a site of consensus and/or reform, but rather an interconnected, heterogeneous, and inherently volatile system to be stabilized or otherwise managed. In this regime, legitimacy is not self-evident so much as it is asserted, produced through, and in turn authorized by, the counterinsurgents' modulation of communication and culture. As such, the core contention of Part Three is that the so-called "safe" generative model embodies and intensifies this logic, being both the product of securitization and a technical agent

of intensified counterinsurgency. This also marks a conceptual departure from output-focused critiques of AI, descriptive accounts of generative media as socially and epistemically corrosive, and critiques of disinformation countermeasures as censorship. Here and elsewhere, I advance an understanding of generative AI as a political technology co-constituting an assemblage of regulatory, technical, and epistemic closures which together have the effect of intensifying and reauthorizing the discursive authority of Visuality 1 in crisis. I contend that these closures do not merely suppress harmful use-cases and content, but instead actively construct the model as a means of strategically *disaggregating* existing and potential insurgency—denying insurgencies of all kinds the aesthetic and technical grounds for their self-visualization by severing the communicative links through which they might coalesce as political subjects and assert the right to look. The core contention with closure, then, is not only that certain information, views, or patterns of interaction are suppressed, but that the infrastructural conditions for insurgent coherence are themselves at risk of being systematically fractured and depoliticized as 'safety.'

In developing this claim, Part Three reviews the safe generative model's positioning as central to infrastructures of culture and communication, and unpacks the regulatory, technical, and epistemic closures constructing it as an instrument of disaggregation. The closures considered in the following pages are not offered as isolated case studies, but as a complex of interrelated and mutually-reinforcing operations, diagnostic of an emergent assemblage of intensified counterinsurgent control.

Positioning the "Safe" Generative Model

Invisuality, Generative AI, and Visuality 1

In *The Vision Machine*, his book on the history of technology and perception, Paul Virilio wrote of an "impending mutation" in the domain of visuality, when the status of human perception as site for the interpretation and management of reality would be supplanted by the emergence of a *sightless vision* and the automation of visualization by what he called the "computerized vision machine."[1] This shift, he noted, threatened an altogether ablation of the human from the work of interpreting and assigning meaning to the world as these processes become increasingly technical, giving rise to a regime of machine visualization wherein instrumental images created "*by the machine for the machine*" form a king of "mechanized imaginary" totally removed from human agency or understanding.[2] Thirty years later in 2024, Virilio's foretelling the rise of a non-human visuality (insofar as it is non-*representational*) reads more like prophecy than prediction. Writing from the other side of its emergence, contemporary artists and theorists have begun to unpack the primacy of these "instrumental virtual images" to the automated operations of a visuality now capable of exercising power at previously impossible scales—from society-wide epistemic management to the incremental modulation of individual citizens through practices of 'nudging' and algorithmic hyper-personalization.

In training the machine learning systems at the center of modern sociotechnical control, human cultural and material life (to the extent that they've been digitized, at least) undergo something of a transmutation, from human-readable representation (in the form of writing, photographs, art, music, etc.) into massive assemblages of unstructured data bearing only an abstract relation to its source.[4] Once trained, the resulting models autonomously "construct diagrammatic abstractions of features common in [the data], and gather these localized

abstractions into predictive statements" which are then operationalized and, sometimes, made legible again in the form of new visualizations.[5] Adrian MacKenzie and Anna Munster have described this intermediary perception as both constituting and taking place within "a field of distributed *in*visuality" in which statistical relationships between data instances come to "count more than any indexicality or iconicity" of the representations which are their source.[6] Quantified so as to be made 'platform-ready,' individual representations thus give way to the non-representational image *ensemble* as the operative unit of vision within what a reformulated visuality the authors call "platform seeing."[8] In effecting this shift from an information and visual culture grounded in human perception to one produced and managed through the operations of platform seeing, technologies of machine learning amplify and re-encode the statistical patterns that are dominant in their training data; that is, they become powerful engines for automating the discursive and epistemological reproduction of Visuality 1 while further denying the right to look through a discorrelation of this work from the mental and perceptual capacities of those subject to its influence.*

We'll remember from Part One that modern visuality—formed through historically-specific complexes of classification, separation, and aestheticization—serves to naturalize particular ways of ordering reality, with the evolution of media technology often serving as a driver of its intensification. Machine learning technologies like predictive analytics, algorithmic recommendation, and now generative AI are no different. In fact, the fundamental principles of visuality closely mirror those of supervised machine learning, where algorithms are trained to

* For more on the effects of this discorrelation of perception on the production of visuality and subjectivity formation see: Shane Denson, *Discorrelated Images*. Durham: Duke U. Press, 2020.

identify and sort new information according to the various sub-classes and patterns of association observable within a corpus of annotated training data.

As these technologies become more prominent in the sociotechnical milieu, so increases the epistemological and material power of the classificatory schemes by which reality is captured, cleaned, and made available for machine learning. Since neural networks do not strictly invent their own classes, the operations of classification and separation at the heart of machine learning necessarily reflect the "historical, geographical, racial, and socio-economic positions of their trainers"[9] as well as the discursive legacies deeply-embedded in Western culture by centuries of plantation, imperial, and now counterinsurgent visuality.[10] In fact, in order to function *as expected*, machine learning algorithms must "assume a world that is regular enough to be entirely described in terms of [the] functional [*invisual*] relationships"[11] circumscribed in their training data. As Kate Crawford observes, given that the datasets and infrastructures of AI often pass as objective technical objects and processes, they tend to stabilize contested categories by producing effects which "are seen to justify their original ordering," even as they encode political assumptions within their taxonomies.[12] And so, it becomes clear how the quickening development and deployment of machine learning serves to embed the epistemological and discursive influence of Visuality 1, due in no small part to the parallels between the series of operations that produces each. The underlying algorithms of machine learning thus not only insert this power "more deeply into the social body"[14] itself, but into the technical and material infrastructures which increasingly govern how we interact and share information, which in turn visualize a reality "in which certain subject positions are made more real and available to us" according to the normative frames which order the data. As such, the array of new technologies

emerging from the field of machine learning have the capability to function as powerful means of social and epistemic regulation while simultaneously obscuring their political nature.

As a subset of machine learning, generative AI share this predilection for reproduction and regulation. Multi-modal models, which are quickly becoming the dominant configuration, are trained on large datasets of mixed media types in order to learn the patterns of association between and across text (in the form of descriptive text and annotation tags), images (in the form of pixel intensities and connections), and temporal media like audio or video (in the form of the probabilistic relationships between one data point and those that come before or after it). The resulting models are then able to generate new text, images, audio, or video (and, increasingly, combinations of each) by generating new data instances which exhibit similar relationships between the different modalities. By way of example, having been trained on many thousands or millions of word/sound pairings, a model optimized for text-to-audio will have established a map of strong statistical associations between particular combinations of characters and the waveforms characteristic of phonetic speech and/or other non-verbal sounds. What results—provided a natural language prompt (for example: “A red cardinal calling from the trees.”)—is a model capable of generating entirely new waveform expressions of the prompt in the form of either synthetic speech, or the sound of a bird’s call. The same principle applies to the way models autonomously make associations across multiple modalities of training data, whether it be text/image pairs, the probabilistic calculation of linear relationships between points in text-only datasets, or virtually any other combination.

While the concept of model weighting* is central to generative AI, these learned associations have effects extending well beyond the pixel-level make up of an eyeball or the likely progression of word strings in response to a prompt. As has been highlighted by a number of commentators (notably Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell in their 2021 paper "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?"), when trained on the internet-based datasets typical to the field of machine learning, language models and other generative AI almost universally re-encode the dominant/hegemonic classifications and correlations inherent to the internet as a source of data which, in English-language contexts, disproportionately leans young, white, and male.[15] The repercussions of this over-representation quickly become apparent when models trained on large internet-based datasets (like Common Crawl,[16] ImageNet,[17] or Flickr-Faces-HQ[18]) are prompted to generate new data with specific reference to categories of class, politics, or race. While not particularly current, the well-known example[19] below offers a clear illustration of this at work:

* In machine learning, *weights* refer to numerical values encoding the strength of association between different features in training data. In a generative model, weights determine how the model maps inputs to patterns it has learned across other modalities—for instance, the association between the text prompt "dog" and visual features such as fur, four legs, a snout, etc. Model training adjusts these weights to reflect how often or closely multimodal features occur together in the semantic space of its training data, enabling the model to generate outputs that are coherent because they reflect those learned correlations.

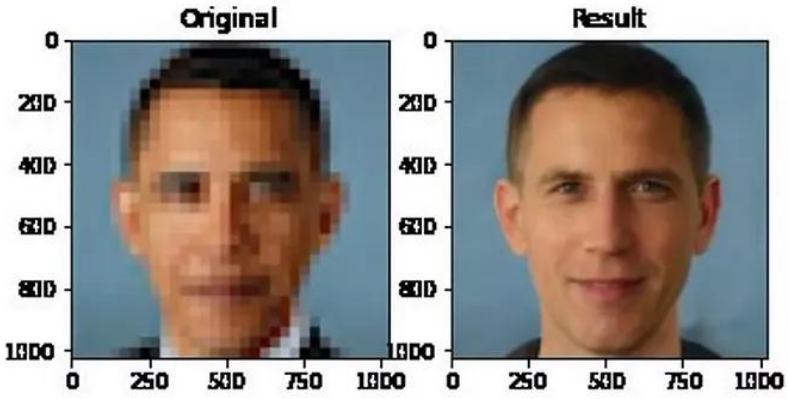


Figure 1 – The PULSE program's upsampling of Barack Obama as a white male.

When tasked with upsampling the pixelated image of a clearly identifiable Barack Obama, the algorithm returns a high-resolution image of a white male. Through what its authors call "latent space exploration," PULSE—the program that created the image above—upscales the low resolution image on the left by traversing " the high-resolution natural image manifold" (or latent space*) of a secondary model—in this case, StyleGAN[20]—in search of images that its algorithm is able to "downscale to the original [low resolution] image."[21] While the PULSE algorithm is clearly very adept at correlating low-resolution data to vectors in the generative model's latent space, it simultaneously reveals the over-representation of specific features in its original dataset. In this instance, scraped from the online photo-sharing website Flickr, the training data preserves

* In machine learning, latent space refers to the body of abstract data that is produced by the training process. As they train, generative models identify statistically meaningful distributions and variances in their training data and "map" them onto the latent space. While not representational in itself, the latent space is a multi-dimensional repository of discrete data points which can then be explored along "vectors" corresponding to features present across the original dataset.

the demographic distributions observable in that service's user base, resulting in a repository of facial images which overwhelmingly skews white.[22] Having been designed to mimic these distributions, both StyleGAN and the PULSE algorithm function to reproduce whiteness by default, and so it stands to reason that—absent intervention—the above results would repeated in all instances except where there is little to no ambiguity as to the race of the person depicted in the original image.

Similar phenomena are readily observable in other generative models, whether it be the over-identification of identity characteristics with certain social attributes by diffusion-based text-to-image generators,[23] or the varied political biases of large language models like BERT, LLaMa and GPT-4 which then "reinforce the polarization present in pretraining corpora."[24] While this has been flagged as an issue by both developers and critics of generative AI, it is important to note that the degree to which a given model reproduces the social and political formations learned in the course of its training is in direct relation to its *success* as a technical process: "Indeed, the reason that bias is framed as a 'problem' is not because the model is making a statistical error, but because it is portraying with *devastating accuracy* the [discursive frames] that historically dominate Western […] culture."[25] Beyond superficial attempts at mitigating biased outputs (discussed later), the efficacy and accuracy of leading generative models continues to be evaluated according to how faithfully they mime and therefore naturalize "our *present* order of knowledge—an order of knowledge that is indispensable to the continued reproduction of our present neoliberal/neo-imperial […] global order of words and of things":[26] Visuality 1. It is also worth noting here that constructing functional models necessarily requires a significant amount of discursive intervention on part of their creators, given the unstructured nature of the internet as a source of data (also discussed later).

It is of course no coincidence that machine learning should exhibit the military objectives of command and control, given that early research into AI was almost exclusively bankrolled by the US Department of Defense through speculative ARPA and DARPA project funding.[27] Reliant as they are on regimes of classification rooted in the historical, geographical, racial, and socio-economic categories of Visuality 1, I contend that the machine learning algorithms behind generative AI perform counterinsurgency as a matter of course. In order to output new information emulating the patterns observable in a given dataset, generative models must "amplify aspects of the input that are important for discrimination and *suppress* irrelevant variations"[28] which fall outside the norm, rendering the statistical outlier (whether cultural, affective, aesthetic, or political) as deviant or irrelevant: "something that can be safely discarded."[29] This emphasis on faithfully reproducing the discursive and epistemological frames native to the world constructed as a dataset is such that the field has invoked the term 'hallucination' to refer to outputs which do not rigidly reflect the models intended classes or weights, despite sometimes representing statistically significant data characteristics.[31] So optimized for reproduction, generative models extend the operations of counterinsurgent visuality within what Louise Amoore has called a new *regime of recognizability*, dictating and policing legitimacy via the automated arbitration of legibility within the "occluded landscape" of data.[32] As we've seen, this arbitration is as much a political process as it is a technical one, and leading generative models produce and enforce a regime of legitimacy which disproportionately affords weight to the formations of Visuality 1 while ranking counter-hegemonic patterns or features as statistical aberrations not to be reproduced. As the operations of machine learning and generative AI increasingly become "the contemporary conditions of recognizability as such,"[34] they partake in circumscribing a restricted horizon of possibility and automate the management

of insurgency as the precondition for visualization in an intensified complex of counterinsurgent control.

It is here that I wish to turn away from counterinsurgent visualization as a political operation inherent to generative models, to begin the work of examining their intentional operationalization as epistemic agents in the sociotechnical context of the global war on disinformation.

Ubiquitous Epistemic Agents

While mainstream generative models are inherently counterinsurgent insofar as they "suppress a particular insurgency [outlier and aberrant patterns of countervisualization] and strengthen the resilience of a particular threatened society or government [Visuality 1 in crisis],"[35] it is important to note that the underlying statistical operations are, on their own, fundamentally value neutral. Being as it is that—when functioning as intended—generative models extend epistemological and representational conventions according to the specific weights assigned during training, they can just as easily 'hallucinate,' misrepresent data, and amplify undesirable biases, or (crucially) be trained to reflect different regimes of historical, racial, and socio-economic legitimacy altogether. Therefore, in the context of crisis and the global war on disinformation (elaborated in previous chapter on the narrative of disinformation/epistemic chaos as emblematic of what Mirzoeff, in 2011, called the crisis of visuality) the technology has overwhelmingly been presented as a challenge to the 'epistemic integrity' of democratic societies and Visuality 1's long-standing monopoly on the means of

visualization, carrying with it the possibility of a dramatic increase in the volume, visibility and verisimilitude of insurgent countervisualization. As such, we are witness to an emergent complex of controls aimed at nullifying the threat of insurgency while further advancing the capabilities of generative AI in favour of a status-quo neoliberal politics and a perpetual re-legitimization of centralized authority through "the seemingly contradictory practice of counterinsurgent governance."[36] So conceived, I contend that generative AI represent a two-fold intensification of counterinsurgency: as they are increasingly embedded across platforms and services, generative AI promise to effectively "automate the visual layer of communication in cybernetic systems"[37] and in turn sharpen relations of epistemic domination on internet; while at the same time an emerging complex of regulatory, technical and epistemic closures participate in their construction as both a discursive and technical apparatus central to a doctrine of counterinsurgent disaggregation.

At the time of writing, trends in the development and promotion of generative AI have led many industry leaders and commentators to predict a near future where generative models form a ubiquitous layer in software stacks ranging from communication and entertainment, to productivity and education. Despite hesitation[38] from some corners, the character of this conversation has generally been dictated by techno-optimists like Jensen Huang, the unsurprisingly bullish CEO of chip-maker NVIDIA, who presented the following vision during his keynote presentation to the Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH) in August 2023:

> For the last 60 years, ever since the IBM System/360, central processing units, or general-purpose computing, was relatively mainstream. […] Well, now, general-purpose computing is going to give way to accelerated computing and AI computing, and let me

> illustrate to you why: *The canonical use-case of the future is a large language model on the front end of just about everything. Every single application, every single database, whenever you interact with a computer, you will likely be first engaging a large language model*.[39]

In a computing future like the one Huang describes, LLMs and other generative models would form a critical part of the information infrastructure, working in concert with existing applications and AI systems to first infer the context and intent of individual queries and then make decisions about how to "present information […] in the best possible way."[40] As such, generative models would take on a rather significant amount of epistemic agency, insofar as they appear poised to further intervene in the processes by which individuals access and interpret information.

While the concept of epistemic agency commonly references the ability (and responsibility) of human agents to form and justify beliefs, the term has more recently been applied to understandings of the "passive or active capacity of a system, organisation [sic], artefact [sic], or technology to impact or influence a person's beliefs and worldview."[41] Advancing this reading, the authors of the 2022 paper "The Internet as an Epistemic Agent" contend that, through becoming powerful epistemic agents in their own right, the technologies which make up what they call the *internet complex* (namely social media platforms and the algorithms which govern them) further our growing displacement as primary agents, increasing in number and complexity the processes of mediation between the source of information and its receiver. When these processes are either hidden or otherwise obscured (as with the black-boxed algorithms governing platform visibility, and now generative AI) users are at risk of becoming secondary epistemic agents as their "contact with reality is mediated by some opaque

interface."[42] Leaving aside the argument that epistemic distance has *always* been assured by one mediating process or another, it is clear that visions of generative AI as epistemic agents primary to an intensified internet complex have vast implications for the future of individual and collective autonomy and the practice of epistemic counterinsurgency online.

In early-2024, one can already see this future taking shape in the form of existing and projected applications/embeddings of generative AI across a number of commercial and consumer-facing products or services: Facebook parent company (and developer of the open-source model LLaMa) Meta is reportedly testing integrations with its social media and messaging platforms allowing users to generate posts, images, and other social content—even "helping you stay active in your communities" by suggesting topics for conversation—as well as modulate the visibility of other user-generated content;[43] Microsoft's recently-launched Copilot integrates OpenAI's GPT-4 across the company's suite of productivity applications, promising a "new *knowledge model* for every organization" from which it will summarize and conduct correspondence, analyze core business metrics, contribute to meetings, and make real-time policy recommendations;[44] And, finally, new processor offerings from Apple[45] and Qualcomm[46] bring the capacity to run generative AI natively on a majority of new smartphone and personal computer hardware.

In addition to the examples above, in mid-2023 Google started beta-testing its Search Generative Experience (SGE). The latest step of a long-running experiment in applying AI to search, the company claims that SGE will "revolutionize how people engage with information."[47] Beyond the promise of providing users a better product experience by replacing traditional Boolean search expressions with natural language prompts, SGE hinges on a new feature Google calls 'snapshots.' Appearing at the top of a search result, these snapshots—

generated by Gemini, their in-house multimodal model—present a text overview or summary of the query or subject matter. While the SGE interface includes citations and references to the source materials used in generating these snapshots (this attribution has also been demonstrated to be unreliable),[48] early analysis suggests that many users of the service will end their search there.[49] Couple this with the added functionality allowing users to generate synthetic media in addition to, or in place of, content already discoverable through search, and the 'SGE experiment' could see Google further cemented as a sole point of contact for those looking to access information online.

This effect is taken to an extreme by the company Perplexity, a start-up whose goals include upending Google's monopoly on search with the introduction of their "answer engine," a generative system which produces standalone text in response to user queries. Speaking to the Wall Street Journal in early 2024, Perplexity CEO Aravind Srinivas said "If you can directly answer somebody's question, nobody needs those 10 blue links"[50]—a direct shot at Google's already-shaky commitment to providing users the context necessary to evaluate the claims made by SGE and other products. The shifting function of search from providing access to information to giving definitive answers points to the increasing epistemic agency afforded to generative models. Whereas the ranked results of traditional search engines allow users to explore and assess a diversity of sources, generative search "moves from presenting the list of web pages *to synthesizing information from them*" in the name of efficiency—a shift that Memon and West argue reduces the depth and variety of information available to users, increases the likelihood of bias, and further positions search companies as the arbiters of truth in an increasingly anarchic information landscape.[51]

As technological assemblages combining generative AI with already well-established algorithms for indexing, ranking, and personalization, products like SGE and Perplexity contribute to the positioning of these models as primary epistemic agents online. Situated as a point of translation between users and the world as data, the generative model is thus in the unique position to naturalize inherently political[53] exercises of discursive control behind the authoritative aesthetics of natural language and UX design.

The potential for generative AI to centralize epistemic and discursive power through the modulation or otherwise limiting of user's ability to access information in an unbiased and horizontal way is made even clearer if we consider insights from a collection of recently leaked Google API documents. Appearing online in May 2024, the documents show code—allegedly related to the company's closely-guarded page-rank search algorithm—containing mechanisms for boosting "authoritative" domains in response to controversial or problematic search queries. Analysts suggest that this, along with other functions gleaned from the leaked data, suggest that Google search has been on, and continues down "an inexorable path toward exclusively ranking and sending traffic to big, powerful [domains] that dominate the web [over] small, independent sites and businesses."[54] This propensity toward centralization, of course, extends to the fine-tuning of the generative models powering products like SGE (which have experienced some issues with hallucination and/or regurgitating nonsense Reddit answers or excerpts from *The Onion*).[55] Echoing Amoore again, while nascent, I believe the integration of generative AI into the infrastructure of internet search serves as a particularly potent example of how these models stand poised to become powerful new arbiters of truth and recognizability online, subtly "shaping the conditions of what can be seen and what is rendered invisible,"[56] and thus capable

of further intensifying a doctrine of counterinsurgency in response to the rhetorical threat of dis- and misinformation.

Foundation Models as Platforms

Recent trends in the rapid development and deployment of generative AI across the internet seem to point to a future where the information and communications ecosystems revolve around a relatively small handful of powerful multimodal models. Commonly referred to as foundation models (FMs), these are very large general-purpose neural networks that, once trained, serve as the "common basis from which many task-specific models are built via adaptation."[57] While there are a number of technical characteristics differentiating FMs from other model architectures, their primary innovation stems from the ability to transfer "knowledge" between tasks (a capability which extends beyond what was included in their original training) and their sheer scale, as made possible through hardware improvements, greater data availability, and the introduction of the Transformer architecture, which "leverages the parallelism of the hardware to train much more expressive models than before."[58] Due to the financial, data, and computational resources involved in developing cutting-edge multimodal models, FMs like OpenAI's GPT-4 or Google's Gemini provide a base atop of which other products and services can be quickly and cost-effectively built. As a result, FMs from a short-list of corporations are being embedded across the internet as large and small companies alike turn to OpenAI, Meta, and Google for the backend of their products and services. By way of example,

the current landing page for GPT-4 on OpenAI's website boasts a list of clients including Dropbox, Duolingo, Morgan Stanley and the Government of Iceland.[59]

Because they display relatively low degrees of specificity in regard to their final use, I contend that FMs resemble, augment, and may potentially replace the platform as the preeminent software foundation of digital communications, all the while extending the platform model of decentralized power rooted in balancing the system's permissive characteristics with an underlying architecture of restriction, censorship, and modulatory control. As Alex Williams reminds us, by virtue of being the foundation upon which other products and services are built, digital platforms are in the unique position of both enabling and simultaneously exerting control over the contingent behaviours made possible within and through them,[60] and foundation models function as platforms in and of themselves as much as they are distinct technologies to be incorporated into the core operations of existing products like Facebook, eBay, Instagram, etc. Like their predecessors, an FM's success as a platform directly results from an ability to balance the relationship between constraint and enablement: maintaining "as far as is possible, a reputation as a neutral broker, the constructor of smoothly operating, apparently impartial space," while obscuring the necessarily restrictive statistical and protocological operations which enable the model to operate in the first place.[61] So long as an FM is able to maintain a sense of this equilibrium, software developers will continue to choose it as the base for their work. Owing to the enormous financial and material resources necessary to achieve this, it seems all but assured that large, privately-owned companies will continue to function as the primary depot for infrastructural FMs, committed to profit through offering what Berry and Stockman have called "cultural-production-as-a-service."[62]

While this would no doubt lead to an increase in the availability of cutting-edge models for individuals and under-resourced organizations, it also raises concerns about homogenization and the risk that biases embedded in a small handful of FMs will be reproduced across a wide range of applications.[63] I argue that we should extend our understanding of this liability beyond the simple dispersion of classificatory and representational biases (again, referring to the model's aptitude for reproducing particular discursive constructions learned through training) to include the dissemination of the largely black-boxed protocological and content controls written in by its creators. When a software developer chooses to license an FM as the platform atop of which to construct a new product or service, they not only inherit the model's innately counterinsurgent tendency to evaluate and intervene in the world according to a predetermined regime of legitimacy, but also the array of passive and active controls which govern the limits of possibility within their application, its environment, and the FM taken together as a heterogeneous system.[65] The reality of this trade-off is that it affords technology companies and their institutional partners an out-sized (and largely invisible) degree of control over the limit and character of meaning and representation in the sociotechnical milieu.

While rationales for baking targeted protocological and content controls into foundation models vary, the common imperative of preventing model misuse in the form of generating dis- and misinformation often veers perilously close to (or is perhaps indistinguishable from) the practice of defining "what constitutes (il)legitimate challenges in political discourse."[66] Paraphrasing Alex Williams on the exploitation of platforms as a fundamental mechanism of control, it is the increasing awareness of generative AI as an technological intensifier of Visuality 1 through reproduction of its classificatory and discursive biases, as well as their enhanced capacity for modulation, that enables—and arguable makes inevitable—its later

operationalization as a political technology: it is through the understanding of how these technologies function to extend the discursive and epistemological conventions of Visuality 1 by generating 'new' information in its image that it then becomes possible to "construct them with deliberate intent, and hence to *strategise around them*."[67] To this point we have outlined how generative AI function as intensifiers of Visuality 1 insofar as they amplify its statistical over-representation in the large datasets commonly used for machine learning. Later in this chapter we will discuss with examples how specific protocological and content controls being built into foundation models in the name of countering bias, fighting disinformation, and increasing model safety factor into generative AI's ongoing construction as a technical and rhetorical agent primary to a strategy of intensified epistemic domination and counterinsurgent disaggregation in response to the crisis of visuality.

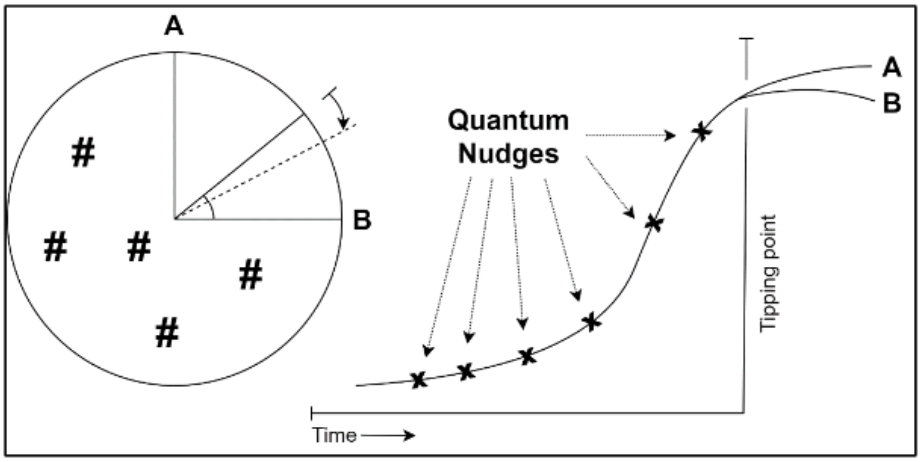
Epistemic Domination

The pending installation of predominantly closed-source generative models at the heart of the communications infrastructure will have wide-reaching effects on the ability of downstream agents to exercise meaningful degrees of epistemic autonomy in their communications and access to information. As ubiquitous agents invested with the social, political and economic interests of well-resourced governments and corporations, foundational models and the array of task-specific applications built atop them represent powerful forces of intensification for conditions of epistemic domination online. A fundamental phenomenon for understanding the political mobilization of communication technologies, epistemic domination has been defined as

"an asymmetrical relation whereby one party has the capacity to control what evidence is available to another, thereby having some degree of control over the latter's ability to acquire knowledge, justified belief, understanding, and related epistemic goods."[68] By directly intervening in the ability of users to exercise a primary epistemic agency, the forecasted ubiquity of proprietary generative AI as mediaries of information and communication represents a barrier to the dominated party's attainment of "epistemically valuable states,"[69] i.e. they stand in the way of an individual or collective's ability to discern and author meaning independent from outside discursive influence—or to assert the right to look. Epistemic domination is thus both an effect of, and prerequisite to strategies of counterinsurgency in the realm of communication and knowledge production.

While conditions of domination obviously predate the internet and generative AI (being a necessary prerequisite to effective propaganda), their ability to automate and naturalize the codification of meaning according to an inherited order of knowledge and protocological control affords their owners new capacities for exercising power over the presentation, withholding or manipulation of the information available in a dominated party's epistemic environment.[70] Beyond the aforementioned examples of Perplexity and Google's SGE (where users' experience of accessing and contextualizing new information is directly modulated by the proprietary code of the service provider), generative AI also enables an unprecedented degree of control over the exchange and character of testimony between members of an epistemically-dominated group. This is enabled through FM integrations with messaging, email and other digital communications in the form of suggested responses, content summaries, and the partial or full automation of

interpersonal correspondence.* In all instances, the epistemic interventions effected by a given model's controls do not need to be overt in order to shape the dominated party's epistemic environment in ways that can influence group dynamics in politically meaningful ways. Researchers affiliated with Canada's Department of National Defense have proposed the term "Quantum Nudges" to refer to the effect of (often minute) discursive interventions which, over time, instill changes in an individual or group's frame of thought regarding a particular subject or situation:



> " [If] an individual is constantly exposed to AI generated content that imperceptibly instill slight changes in the frame of thought, then the opinion of an audience could be slowly brought to the tipping point depicted in the above figure. If A and B are two meanings for a hashtag, […] the likelihood of meaning B is reinforced past the tipping point by this process."[71]

Figure 2 – Illustration of public opinion manipulation through the cumulative effect of quantum nudges.

---

* In December 2023, Meta introduced multimodal capabilities to its Messenger application (https://about.fb.com/news/2023/12/meta-ai-updates/). [accessed February 28, 2024] Also in late 2023, Microsoft introduced Copilot for Outlook, capable of summarizing long email conversations and generating potential replies (https://support.microsoft.com/en-us/copilot-outlook)

[accessed February 28, 2024]

While the example above specifically references generative AI in the context of organized influence operations and their cumulative effects on information environment assessment and cognitive security at the population level, the concept of the Quantum Nudge can as easily be applied to understanding the operationalization of FMs as a political technology and the imposition of protocological and content controls—"structuring a negotiated dominance of certain flows over other flows"[72] —as methods for restricting or modulating the ability of users to freely generate, receive, and communicate information.

The concept of "nudging" as a means of social engineering has been variously explored, and has recently reemerged in popular conversation on the increasingly visible politics of AI systems. For example, in an article published by *The Atlantic* in March 2024, the author cautions readers against succumbing to the "soft mind control" of AI nudging, citing an exchange with Google's Gemini wherein the chatbot generated campaign materials promoting vegetarianism after having refused earlier prompts regarding a meat-based diet.[73] While this example is relatively innocuous, the author goes on to describe similar experiences regarding more politically and socially-charged prompts, with Gemini subtly nudging the user away from topics and positions the model determines to be uncouth. However, as illuminated by critical disinformation scholarship, the evidentiary basis for asserting a causal relationship between such nudges and changes in political or social attitudes remains tenuous at best—often inferred rather than demonstrated, and more reflective of elite anxieties than substantiated effects. Still, even in the absence of clear causal influence, the principle of nudging remains analytically useful for understanding how generative models participate in the disaggregation of insurgency—subtly

redirecting attention, foreclosing certain lines of inquiry or connection, and undermining organization.

In the case of leading foundation models, the negotiation of information flows tends to take shape around a nebulous concept of protecting and promoting "safe" discourse across the platforms which employ that particular model. Of course, determining what does and does not constitute safe in this context is a socially- and political-fraught endeavor, and so these determinations and the resulting controls (through enacting degrees of modulation or outright refusal) take part in shaping generative AI as a means for nudging the content and character of online communications, managing the terms of visibility, and shoring up Visuality 1 against a growing tide of "unsafe" insurgent visualization. As will be discussed later in this chapter, while they identify and reject undeniably corrosive representations around subjects of race, gender and identity, the array of currently proposed closures rigidly align generative AI with a worldview based around otherwise status-quo formations of discursive, sociopolitical and economic power, and therefore inform the installation of controls which threaten to similarly inhibit the expression of legitimate dissent and otherwise possibility.

Disaggregation

To discuss generative AI's political mobilization in service of a doctrine of counterinsurgency, I should note how this framing differs from those which would consider the technology as simply extending a paradigm of propaganda that came to dominance alongside the mass media of the twentieth century. As Jacques Ellul described it, modern propaganda operates

by constructing an all-encompassing (and ultimately closed) system of meaning through which subjects come to know and act in the world. Through patient and repeated exposure, this "organized myth," as Ellul called it, cements its own legitimacy and eventually "takes possession of a man's [sic] mind," having become the reflexive framework through which all experience is interpreted.[74] This is total propaganda as *mythopoesis*, a campaign of mass visualization which conscripts media in the arousal of an "active and mythical belief" in the totality and legitimacy of its authority,[76] thus persuading its receiver to think and act accordingly.

While generative AI displays mythopoetic characteristics insofar as it participates in economizing aesthetic production, I would argue (in line with many critical disinformation scholars) that it does not do so in a way that is any more convincing or effective than the media which preceded it. What makes generative AI unique as tool of propaganda is not its capacity to foster stability through the aestheticization and legitimation of myth, but its utility as a tool for *preserving myth* against challenges to its authority, maintaining the appearance of totality, stability, and cohesion on the belief that legitimacy will follow. In this sense, it emerges not as the herald of a renewed ideological project, but as a defensive apparatus deployed amid the crisis of visuality—an attempt to hold together an already fracturing regime of legitimacy. Unlike the propaganda of prior eras, counterinsurgent media operate without conviction, in a political environment where the establishment has no myth left to promote, only threats to suppress.

As developed and deployed in the context of the global war on disinformation, I intend to illustrate that generative AI is capable of automating the primary discursive activities of counterinsurgent visuality by installing an infrastructure which visualizes the social as bifurcated along an axis of legitimate and illegitimate patterns of information and interaction, and in turn partakes in the tactical barring or excision of countervisualization from the mainstream epistemic

field. While the strategic aims of classic counterinsurgency ultimately include political resolution, "*tactical* actions to counter an insurgency primarily serve the purpose of buying time for the political solution to be implemented."[78] However, as we've seen, in an environment dominated by a capitalist realist politics of neoliberal inevitability, the instantiation of a border between the insurgency and 'host' culture comes to stand as an end in-and-of itself, creating enough 'security' so that business can proceed as usual. Framed by the rhetorical threat of disinformation and infocalypse, generative AI thus becomes both problem and solution to the rise of increasingly popular insurgencies on the platform internet: by elevating the issue of countervisualization to the level of 'existential threat,' the technology simultaneously serves to justify the authoring and imposition of extraordinary information controls through a complex of regulatory, technical and epistemic closures which together shape it as a powerful agent of counterinsurgent intensification.

In the context of the historical struggle between Visuality 1 and its counters, we can define insurgency as the organized practice of visualizing reality in ways that are, in a literal sense, "against the rules" established by and in favour of dominant visuality.[79] In the parlance of this dissertation, these "rules" refer of course to the discursive (and in turn, material) operations of classification, separation, and aestheticization which make up the historically-specific complexes of Visuality 1. Central to the establishment and practice of one party's epistemic domination of another, rules dictating the terms by which reality is to be ordered and understood serve to first delineate a dominant regime of legitimacy or truth and, as a corollary, perform the counterinsurgent function of delegitimizing *in advance* the discursive formations and visualizations of groups denied entry to that regime. As such, insurgency refers to those

epistemologies and rhetorics which, under conditions of opposed or contested governance, operate *from the outside* to unsettle a hegemonic order.

As discussed in the previous chapter, the label of disinformation has emerged as the primary device for naming the phenomena of insurgent challenges to the legitimacy of a complex that has for the better part of a century leaned centrist, liberal, and Western. While the term is most commonly evoked in reference to information operations coordinated by foreign or domestic political adversaries with illiberal tendencies, I contend that the range of measures taken in response threaten to similarly enable the suppression of more or less legitimate social and political dissensus. In analyzing generative AI's operationalization in service of counterinsurgency, my following analysis works from a definition of insurgency that takes disinformation to be inclusive of all visualization practices which function to, directly or indirectly, destabilize the discursive hegemony of establishment visuality.

If insurgency functions to disrupt hegemony, the aim of counterinsurgency, then, is to return the host society to a "stable, peaceful mode of interaction—on terms favorable to the government."[80] When thinking FMs as control systems, then, counterinsurgent control can be understood as what is enacted through the network of homeostatic feedback mechanisms which intervene in the system in order to "maintain a target state"[81] of functionality and safety. Of course, we have already discussed generative models insofar as their basic functioning relies on analyzing user prompts and apportioning positive and negative feedback according to the statistical patterns observed in their training data, thereby enacting a degree of counterinsurgency that precedes political interventions introduced later in the form of tailored protocological and content-based controls, or in the industry parlance, *alignment*. Later in this chapter, I examine how the emerging complex of regulatory, technical and epistemic measures taken in response to

generative AI as an existential threat together apply negative feedback in what I argue amounts to the technology's systemic capture in service of a doctrine of counterinsurgency.

Driven by the threat of 'infocalypse' or epistemic chaos, this strategy employs generative AI not simply as technical means of epistemic modulation or suppression, but as a discursive device for justifying a progressive intensification of counterinsurgent control under the imperative 'society must be defended.' As such, the securitizing narratives of disinformation and existential risk thus reflect one of the central tenets of counterinsurgent doctrine: the mobilization of popular support through the exploitation of an alternative narrative which excludes the insurgent and "emphasizes the inevitability and rightness of your ultimate success."[83] In this framework, popular support for the counterinsurgent is less grounded in political or ideological agreement than it is in the perception of legitimacy and a belief that dominant institutions can ensure security and act in the public interest. As David Kilcullen puts it, "For your side to win, the people do not have to like you, but they must respect you, accept that your actions benefit them, and trust your integrity and ability to deliver on promises, *particularly regarding their security*."[84]

Once a narrative has been established yoking stability to the dominant regime, incumbent powers are able to present the insurgency as fundamentally *destabilizing* (even when this is simultaneously its goal) and therefore contrary to the population's interests, thus allowing even the most direct and/or unpopular applications of force to be presented as both measured and necessary. By raising the spectre of irreversible epistemic decay and ensuing democratic collapse, the narrative of disinformation functions to buy increasingly unpopular institutions of neoliberal governance the political goodwill necessary to impose powerful new control measures under the banner of 'maintaining information integrity,' all the while exonerating them for their

political economic stake in creating and sustaining the structural and epistemic inequities which produce insurgency in the first place.

Working in concert with and extending the existing counterinsurgent infrastructures of algorithmic control and surveillance, generative AI intensifies counterinsurgent visuality by providing the dominant regime additional means to visualize and distinguish the insurgent, and in turn nullify it through precise applications of force in the form of regulation, discursive modulation and/or censorship, and various epistemic defeaters. I propose that this particular faculty can be understood as proceeding according to a strategy of counterinsurgency that military scholars call *disaggregation*. Tailored to the era of globalization, strategies of disaggregation are central to a model of counterinsurgency, underpinned by complex systems thinking, which understands insurgencies as "social systems" formed when preexisting elements of a society "organize themselves in new patterns of interaction involving rebellion, terrorism, and other violent political activity."[85] In the systems model, the threat of insurgency lies not in its individual elements (disillusionment, radicalized individuals, material supports) but in the *patterns of interaction* between them. Understood this way, insurgency only rises to the level of threat when previously-atomized threads of dissent are able to connect, or aggregate, along "energy pathways that allow disparate groups to function in an aggregated fashion across intercontinental distances" and thus become capable of functioning as a global system of resistance.[87]

Emerging from a theoretical recalibration of what constitutes 'success' in global counterinsurgency (prompted by early failures in countering 'global jihad' in the theatres of Iraq and Afghanistan), strategies of disaggregation reflect the prioritization of tactical disruptions to the insurgency's ability to function as a system over the stickier and more ambiguous objectives

of sociopolitical amelioration and the eventual reformation of adherents to insurgent ideologies. Because this framework sees insurgency first and foremost as an undesirable pattern of interactions between previously disparate elements, disaggregation shifts the strategic goals of counterinsurgency from eliminating or reforming individual elements, to developing methods for forcibly returning the system to a "normal" mode of interaction between elements. I think it's particularly interesting to note that disaggregation may be read as a primary discursive operation of visuality itself, insofar as the complex Mirzoeff calls Visuality 1 sustains its authority through the separation of those it visualizes in order to "*prevent them from cohering as political subjects*" capable of contesting its legitimacy.[88] While disaggregation is observable in the operations of each complex of visuality, I contend that its present intensification in the form of the 'safe' generative model proceeds through a number of methods specific to modern counterinsurgency.

In Kilcullen's formulation of disaggregation, there are a limited number of ways to disrupt insurgent systems: "One can (1) attack nodes, (2) interdict links, (3) disrupt the boundary, (4) suppress boundary interactions, (5) choke off inputs, (6) deny outputs, or (7) use a combination of methods."[89] Of these, I argue that closures shaping generative AI as an means of disaggregation focus primarily on its utility to the following modes of 'attack': interdicting links between local insurgencies and global networks of ideologically-aligned actors by disrupting or infiltrating lines of communication; denying and/or modulating the output of new grievances and propaganda by restricting or controlling the insurgency's access to its population via mainstream platforms; and choking off inputs in the form of the human, material and informational resources necessary to sustain the insurgency under changing technological and political conditions. The theory behind disaggregation holds that, taken together, these modes of attack function to effectively asphyxiate insurgency, cutting it off from itself and the population, and forcibly

rearranging disordered patterns of interaction back into "a stable and peaceful 'system state'" organized on terms favorable to the counterinsurgent regime.[90]

In the pages to follow, I will present a representative selection of regulatory, technical, and epistemic measures proposed in response to the potentially destabilizing effects of generative AI. Following an account of their properties, the measures will then be analyzed as co-constitutive elements of a strategy of disaggregation carried out against the tide of insurgent countervisualization colloquially referred to under the umbrellas of 'disinformation' and 'crisis.' According to the proposed framework, *regulatory measures* function to choke off inputs by restricting access to the resources and materials necessary to develop and deploy powerful generative models; *technical measures* interdict links and deny outputs through the imposition of targeted protocological and content controls intended to modulate or restrict "unsafe" visualization; and *epistemic measures* which—in the form of labeling, authenticity and provenance initiatives—reinforce conditions of epistemic domination, exploiting a single narrative that "excludes the insurgents"[91] and thus "defeats"[92] the epistemic status of countervisualization by reasserting the authority of traditional gatekeepers as the arbiters of truth in an increasingly anarchic media landscape.

Closure

After the preceding pages' account of the generative model's positioning as a central mediator of communication and knowledge, this next section turns to the specific regulatory, technical, and epistemic closures attempting to render it functionally "safe," marking a shift from

theorizing the model's epistemic and political function to examining the mechanisms by which that function is being secured. The following analysis traces how these controls thus work in concert to render the model—and its epistemic environment—"safe." In this reading, the emergent complex of closure is more than a collection of reactive safeguards against malign interference, but rather constitutes intervention: depoliticizing the model's consolidation as an instrument of counterinsurgent disaggregation.

Regulatory Closure

As generative models increasingly find application in products and information infrastructures ranging from social media and private communications to policy analysis and national defense, there has unsurprisingly been a rush to develop regulation tailored to addressing the specific challenges they represent. Routinely citing the need to assess and mitigate the technology's inherent risks (the most commonly specified of which include bias and discrimination, fraud, data privacy concerns, informational hazards, and—of course—disinformation) a number of draft regulations have recently[*] been put forth by government and corporate bodies around the world. Of the many proposed, this section of the dissertation will focus on an analysis of the standards set forth in acts of legislation and codes of conduct currently taking shape in the European Union, People's Republic of China, and the United States of America.[†] To the extent possible, this analysis will focus on these regulation's specific

[*] This chapter was first drafted in February and March of 2024.

[†] At the time of writing, Canada does not have formal regulation in place that is specific to AI. In September 2023, the federal government issued the temporary "Voluntary Code of Conduct on the Responsible

treatment or implications for the development and deployment of foundation models and downstream products or services. While not examined in detail, earlier multilateral initiatives such as the Hiroshima Process[93] and the Bletchley Summit[94] are worth noting as discursive precursors—helping to consolidate a shared vocabulary of "AI safety" and establish a trajectory for the formal regulatory frameworks now emerging. Following a brief summarization of their content and character, I consider how the frameworks' invocation of the ambiguous and politically-malleable concept of 'AI safety' as a guiding principle for regulation threatens to produce a global standard of governance concentrating power in the hands of sanctioned developers, thus internalizing the logic of counterinsurgency by denying inputs to producers of open-source or otherwise unsanctioned or 'unsafe' foundation models.

European Union: The Artificial Intelligence Act

First introduced in 2021 and passed by parliament in March 2024, the European Union's Artificial Intelligence Act (EU AI Act) is touted as the world's first body of AI-specific regulation to be passed into law.[*] In an effort to introduce regulation that effectively addresses and mitigates liability while still encouraging innovation, the act sets out a four-tiered framework which applies rules to AI systems based on their assessed level of risk. Systems of 'minimal' or 'limited' risk (such as spam filters, e-commerce recommendations, customer service chatbots,

---

Development and Management of Advanced Generative AI Systems," and the Artificial Intelligence and Data Act of 2023 has set out a framework for future legislation.

[*] This despite Chinese regulations taking effect in 2021 and 2022, with draft rules for generative AI expected to come into effect mid-2024. More on this and the importance of precedence later.

etc.) are either unregulated under the current legislation or subject to basic transparency obligations, requiring, for example, that users be informed when they are interacting with an AI system. 'High' risk systems (those uses impacting rights or safety, such as critical infrastructure, employment, or policing) are the focus of the act, and thus subject to extensive regulatory oversight. Finally, 'unacceptable risk' includes applications deemed to pose clear threats to safety, stability, or fundamental rights, and are therefore prohibited entirely.[95]

In the act, risk-level is generally correlated to increases in model size/capability, and developers are required to notify regulators of their intent to build models exceeding a certain threshold. While, according to the scale above, foundation models (referred to in the act as "General Purpose" or GPAI) are identified as being high-risk, EU policymakers have subjected them to additional constraints when they are deemed to pose 'systemic risk.' The act applies this designation to models exhibiting the ability to effect critical sectors such as public health and safety, disrupt democratic and economic processes, or create and disseminate "illegal, false, or discriminatory content."[97] Under these rules, *all* providers of GPAI models are subject to reporting requirements (details about model architecture, training data, and testing protocols) and must demonstrate commitments to respecting EU copyright law. However, models deemed to pose *systemic* risks are subject to additional rules, including independent model evaluation, risk assessments and mitigation measures, cybersecurity protections, and the obligation to report incidents to the newly-formed "AI Office" and other relevant authorities.[99]

As outlined in the act, compliance is expected to be achieved through codes of practice jointly developed by industry and a newly established European Artificial Intelligence Board. Made up of representatives from EU member states, the AI Board, together with an expert advisory panel, is tasked with ensuring the proper implementation and enforcement of the

regulations contained in the act, as well as creating and keeping up-to-date the technical specifications and standards necessary to this work.[100] In June 2025, the European Commission came under pressure from tech companies and some government officials to "stop the clock" on implementing the Act, claiming that the regulatory oversight and codes of practice required under the new rules would harm the bloc's ability to compete in the global market for AI models and services.[101] In July, the Commission confirmed that the Act would proceed according to its original timeline, publishing the final version of a *General-Purpose AICode of Practice* to take effect August 2nd, 2025.[102]

People's Republic of China: Interim Measures for Generative Artificial Intelligence

Contrary to the EU's claim at being the 'first' to introduce regulations tailored to artificial intelligence, AI has long been a legislative priority for the Chinese government, with state investment in the industry identified as key to economic growth and global competition in each of the CCP's most recent Five Year Plans[103], and the 2017 introduction of the *Next Generation Artificial Intelligence Development Plan*.[104] In terms of regulation, current laws addressing generative AI follow the prior introduction of governance frameworks specifically tailored to recommendation algorithms (the *Internet Information Service Algorithmic Recommendation Management Provisions*[105] became law in March 2022) and deepfakes (*Internet Service Deep Synthesis Management Provisions*,[106] January 2023).

In August 2023, the *Interim Measures for Generative Artificial Intelligence Service Management*[107] (which functionally extend the measures laid out in previous regulations to more

pointedly address the evolving state of generative AI) came into effect, becoming the world's first binding national regulations governing generative AI models. Article 1 of the new law lays out its scope and objective, defining the regulations as temporary measures put in place "to promote the healthy development and standardized application of generative artificial intelligence," while also protecting national interests and maintaining public order.[108] While this general framing largely reflects the same motivating principles underpinning the EU AI Act and other international proposals, the Chinese regulations contain notable differences in emphasis—particularly regarding the political function of generative AI. For example, Article 4(1) of the *Interim Measures for Generative Artificial Intelligence Service Management* stipulates that generative systems developed and/or deployed in China must adhere to "core socialist values," prohibiting systems from producing content that may "incite subversion of state power," "endanger national security," "damage the national image," or otherwise threaten social stability through content deemed false or harmful.[109] These prohibitions appear consistently across the country's recent AI governance documents and, when read from a liberal democratic perspective, they underscore the political commitments and motivations behind legislative determinations of 'risk' and 'safety.'

The remainder of the document again resembles some of the boilerplate recommendations shared with other regulatory frameworks, requiring that developers register foundation models with a central agency and have processes in place to quickly alter model behaviour in response to government request.[111] Additionally, draft and interim measures contain orders directing the providers of generative services to collect and retain the "true identity information" of users,[112] label AI-generated content, and submit to regular security audits in accordance with the country's aptly-named "Regulations on Security Assessment of

Internet Information Services with *Public Opinion Attributes or Social Mobilization Ability*"—emphasis added. Following this, it should be unsurprising that, under the new regulations, foreign-developed models are subject to increased scrutiny, with provisions put in place to monitor their behaviour and "employ technical measures" to bring offending systems into compliance.[113] As with the EU rules (and as their name suggests) the interim measures leave ample space for further refinement as the industry develops, with a more detailed regulatory framework expected in the coming years.

United States: Biden's Executive Order 14110, the Decentralized Approach, and Trump's 2025 'AI Action Plan'

Released by the Biden Administration on October 30, 2023, Executive Order 14110 on *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* establishes a government-wide initiative for encouraging innovation and leadership in the "responsible" development of artificial intelligence.[114] Despite outlining an array of policy goals, the order itself is not a piece of standalone regulation, and instead confirms the government's commitment to a decentralized approach to regulation. Characterizing this task as a "society-wide effort that includes government, the private sector, academia, and civil society," the order directs independent entities of the federal government to take actions pursuant to eight "guiding principles and priorities," summarized below:

1. Artificial Intelligence must be safe and secure.

2. The United States must lead the world in AI through promotion of responsible innovation, competition, and collaboration.

3. AI regulation and development must reflect a commitment to American workers.

4. AI policies must respect equity and civil rights.

5. AI products and policies must respect consumer protections.

6. AI policies must protect the privacy and civil liberties of Americans.

7. Manage risks associated with government use of AI systems.

8. The United States must lead the world now as it has through previous eras of disruption.

Further to the objective of safety and security, the order calls for the development of "robust, reliable, repeatable, and standardized evaluations" of AI systems, including national security assessments, "post-deployment performance monitoring," and ensuring resiliency against "misuse or dangerous modifications."[115] Central to this effort is the practice colloquially referred to as 'red-teaming,' a structured form of adversarial testing where researchers try to expose security vulnerabilities, discover dangerous or emergent model behaviours, and otherwise assess the system's susceptibility to misuse.[116] According to the order, AI systems (particularly "dual-use foundation models") must be subject to red-team protocols developed in collaboration with offices of the Department of Homeland Security before being deployed in any systemic capacity.[118] Similar to the EU regulations, the order also suggests reporting requirements and other restrictions for foundation models meeting certain technical conditions, largely to do with some combination of computing power and size of the trained model.[119]

On February 8, 2024, in response to this and other directives contained in the executive order, the National Institute of Standards and Technology (NIST) announced the formation of the US Artificial Intelligence Safety Institute, billing it as a consortium bringing together "more than

200 organizations to develop science-based and empirically backed guidelines and standards for AI measurement and policy, laying the foundation for AI safety across the world."[120] Members include nearly every major US-based technology company alongside organizations from civil society, journalism, finance, and defense (Booz Allen Hamilton, Lockheed Martin, Northropp Grumman, Palantir, RAND Corporation, and SRI International). To date, responses to the executive order have largely taken the form of revamped internal policies at companies developing FMs, and on-the-ground legislative efforts on part of federal agencies like the FCC, FEC, and FTC,[121] as well as individual states.[122]

More recently, the second Trump Administration has signalled a shift away from "safety" as a core motivating principle for AI policy and regulation, rescinding Executive Order 14110 with the signing of Executive Order 14179 "Removing Barriers to American Leadership in Artificial Intelligence," followed by the introduction of EO 14192 "Unleashing Prosperity Through Deregulation," and its *AI Action Plan* in July 2025. Subtitled "Winning the Race," the document lays out a tripartite plan to "achieve and maintain unquestioned and unchallenged global technological dominance" by removing barriers of "red tape and onerous regulation"; rapidly expanding domestic AI development, manufacturing, and infrastructure; and aggressively exporting the full stack of American-origin technologies and standards with the intent of countering foreign influence and positioning the United States at the helm of global AI development, regulation, and use.[123] Framing global AI dominance as critical to the country's geopolitical strategy, national security, and economic prosperity, the act trades the prior administrations language of 'safe and responsible' AI development for the goal of producing technologies and applications that are "secure-by-design," a concept emphasizing interpretability, control, and robustness against adversarial attack over harms-based safety

research—which is unambiguously presented as a leftist "social engineering agenda" standing in the way of rapid innovation.[124]

Regulation and AI Safety/Security Discourse

A survey of the major regulations being enacted or proposed around the world reveals a shared emphasis on ensuring that AI systems be developed and utilized in ways that meet some established standard of "safety." Understood in relation to Mirzoeff's formulation of visuality and the intensifying force of new media, regulations around frontier AI (and FMs in particular) must be read as discourses in and of themselves, with the rhetorical role of 'safety' then interrogated as to who or what interests it serves to protect, from what, and to what ends. As Barry Friedman has argued, the concept of "public safety" often functions as a political project that reflects and reinforces dominant interests—privileging protection from violent threat while sidelining other forms of harm stemming from poor conditions. This framing collapses safety into protection and protection into policing, a narrowing that travels easily into AI governance, where similarly imprecise invocations of "safety" can obscure the political work that these regulations and countermeasures perform.[125]

Deployed in the context of AI regulation, commitments to safety entail the identification and reduction of a model's inherent potential for "harm" (the psychological, ethical, or sociopolitical damages resulting from use), which in turn follow from and compound its "risk" (as the potential for present and future harms).[126] Regulations aimed at promoting AI safety, then, generally focus on evaluating and mitigating risk in order to prevent its realization in the

form of specific harms. While (and perhaps because) the potential harms of 'uncontrolled AI' run the spectrum, typical legislative verbiage around the concepts of AI safety and risk remains imprecise. For example, take this passage from the EU AI Act, which sets out to define 'risk' as a structuring concept of its framework: "[Risk refers to the] actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain."[127] While the concept of safety is relatively unambiguous in relation to AI risks like mass unemployment, impersonation and harassment, data privacy, or mass extinction, when discussing the regulation of FMs insofar as they might cause harm in their application as *epistemic* agents, there is something distinctly political about the task of determining exactly what constitutes risk in regards to public safety, security, and *society as a whole*. In the context of our analysis, 'safe AI,' as a regulatory objective, might then be a read as a state consistent with the re-legitimation of Visuality 1's discursive authority as expressed under the similarly vague imperative of maintaining 'epistemic integrity,' with 'unsafe' generative AI thus standing in as a risk vector for potential harms in the form of disinformation and the 'informational hazard.' As with the term safety, here there is also a certain flexibility afforded by imprecise definition.

In a 2018 survey, researchers for the Hewlett Foundation cited a number of shortcomings inherent to the new field of disinformation studies. Chief among these was a general lack of common definitions for terms like 'hyper-partisan' or 'fake' news, 'conspiracy theory,' and, critically, 'disinformation' itself.[128] The difficulty of formulating a definition suitable to anchor a new field of 'objective' knowledge production perhaps stems from the fact that the term's meanings are so often inextricable from the specific social and political contexts of their application, which are, by their nature, also unstable. For example, the widely-accepted origin of

disinformation's modern usage comes from an entry for *dezinformatsiya* in the 1952 volume of the *Great Soviet Encyclopedia*, which describes it as false information spread by the "capitalist press and radio" for the purposes of misleading public opinion (this was, of course, concurrent with the ongoing operationalization of disinformation as a tool under *active* measures, the USSR's strategy of foreign and domestic political warfare).[129] As such, any practical regulatory definition of disinformation would have to either acknowledge or resolve the dissonance inherent to identifying disinformation as an outside threat while simultaneously employing it as a means of maintaining and exercising political authority through the regulation of discourse. Barring this, the political nature of determinations about "safety" with regards to information technologies like generative AI can thus be found in the language of the regulations themselves, whether it be the Chinese measures' emphasis on protecting and promoting "core socialist values," or the EU (and now to a lesser extent, US) documents' conflation of the neoliberal democratic order with reality such that it needn't be explicitly named.

So, it follows that the implicit meaning of "safety" and "risk" vary according to the social, political and economic commitments of the stakeholders involved in defining and authoring new regulations aimed at curtailing the 'existential risks' of generative AI and disinformation. In an article for *Harpers* on the rise of what he calls "Big Disinfo," Joseph Bernstein has observed that disinformation studies (as the primary stream of research tasked with making these determinations) often reflects the political economic positions of the elite institutions that house and fund them, like the New York Times, Standford University, the Council on Foreign Relations, or the World Economic Forum: "That the most prestigious liberal institutions of the pre-digital age are the most invested in fighting disinformation," he writes, "reveals a lot about what they stand to lose, or hope to regain."[130]

Considered in light of the crisis of visuality (and therefore sidelining more hysterical prognostications of mass death at the hands of malicious AI) the epistemic risks posed by uncontrolled generative AI, while undoubtedly disruptive, become *existential* only insofar as they harbour the potential to further unsettle the already contested salience of Visuality 1 and its organization of power and privilege. As such, "safety" as a goal of governance cannot be understood apart from a consideration of the various interests of those parties invested in the regulatory effort. In this case, the concept of AI 'safety' thus serves to justify the legislative imposition of any number of regulatory, technical and epistemic measures aimed at nullifying the threat of insurgent countervisualization under the auspices of epistemic integrity, returning the disordered patterns of an information ecosystem 'in decay' to "a stable and peaceful 'system state'" organized on terms favorable to an incumbent regime of legitimacy.[132]

This point is particularly salient in light of recent political and discursive changes with regards to how some countries frame artificial intelligence as a matter of concern. In February 2025, the government of the United Kingdom renamed its AI Safety Institute to the 'AI Security Institute,' signalling a shift in emphasis away from research curtailing bias, discrimination, and societal harms and toward "strengthening protections against the risks AI poses to national security and crime."[133] The move has largely been seen as an attempt to align UK AI policy with the priorities of the United States under the second Trump Administration, which in June rebranded its own AI Safety Institute as the 'U.S. Center for AI Standards and Innovation,' now focused on promoting AI development for commercial and national security applications.[134] While both changes effectively erase terms like 'safety' and 'harm' from the political AI lexicon of their respective country's, the trading of 'safety' for 'security' helps to reveal a limited distinction between the concepts in terms of their political and technical effects, as the relaxing

of ‘safety’ regulations makes way for increased cooperation between tech companies and the national security and intelligence communities.

The authors of a 1992 paper on the meaning of safety and security with respect to the design of computer systems make the observation that English is one of the few languages to regard the two terms as distinct (i.e. where as the German *sicherheit* or Chinese *ānquán* mean both, the common English phrase ‘safe and secure’ suggests that they represent different properties).[135] The authors go on to unpack this differentiation by evaluating the two on the basis of causality, defining issues of safety as those involving immediate, *direct* harm, and security as those having to do with enabling or increasing the *potential* for harm—arriving at a distinction reflected in the “basic intuition” that “a system is not safe if it can harm us; it is not secure if it gives other the means of harming us.”[136] According to this definition, with respect to their deployment as information and communication technologies, the functional capacities of generative systems for direct and potential harm (however defined) become indistinguishable insofar as their safety is necessarily contingent on their being made secure against a particular type of use. That is, it is through the same (or at the very least, complementary) mechanisms that the model is rendered both controllable in terms of behaviour and robust against adversarial attack or misuse: a model that does not produce ‘harmful’ content is only *safe* so long as it is also *secure* against its intentional generation.

It is important to note, however, that this does not operate in both directions, and a generative system’s security is *not* contingent on its meeting any particular safety criteria, the definition of which being fundamentally political and thus subject to change. In this way, concerns over safety and the information controls they produce have *always* also been about security, and the decision to replace one term with the other ultimately reveals less about policy than it does about the

specific political values implied by each. As safety loses the currency necessary to legitimize new controls, the turn toward 'AI security' in the US, UK, and elsewhere represents the seemingly inevitable endpoint of the technology's securitization, providing new political and discursive justification for increasingly blatant intensifications of counterinsurgent control.

However, it is important not to attribute this particular stage of intensification strictly to those actors who initiate the aforementioned discursive shift. While safety and security measures create the means to disrupt the information operations of both foreign state actors and domestic threats variously figured as right-wing militias, 'incels,' the 'radical left' or other 'extremist' groups, the resulting counterinsurgent apparatus is effectively neutral, and thus casts a wide net when deployed in defense of varying political regimes. For example, a 2021 document outlining the United States' domestic counter-terrorism strategy describes a growing internal threat from sources broadly defined as "any individual or group [which engages in] opposition to legislative, regulatory, or other actions taken by the government," including but not limited to "environmental" and "animal rights" activists, or simply those who oppose "capitalism, corporate globalization, and governing institutions which they perceive as harmful to society."[137] Similarly, despite being initiated during the first Trump presidency, policy suggestions contained in the 2021 report of the National Security Commission on Artificial Intelligence both informed the Biden administrations policy development regarding AI for national security and intelligence purposes, and would later form the basis of the 2025 AI Action Plan. Perhaps unsurprisingly, the development and deployment of an intensified counterinsurgent apparatus with AI systems at its centre remains constant despite the changing of the guard.

More recently, the demonization of "radical left" and even moderate conservative politicians, media figures, and citizen activists by US president Donald Trump not only expands

the criteria for identifying internal threats and confers the state greater latitude in its deployment of force, but also points to an underlying operational parity between the counterinsurgent goals of this administration and those preceding it. In a statement announcing the agency's reformation, US Secretary of Commerce Howard Lutnick criticized the Biden-era AI Safety Institute as a vehicle for the development of "censorship and regulations […] under the guise of national security," almost certainly referring to debiasing, content moderation, and other 'woke' safety measures that his office claims stand in the way of innovation.[138] However, later in the statement, Lutnick says that the new agency will foster the deployment of advanced AI systems while still "ensuring they remain secure to our national security standards"—all but confirming that this derision does not extend to the likewise regulation, surveillance, and censorship of 'terrorist' or 'extremist' threats as defined by the current administration. That the current State Department's AI-driven "Catch and Revoke"[139] initiative—authorized by EO 14161 "Protecting the United States from Foreign Terrorists and Other National Security and Public Safety Threats"—utilizes a digital surveillance infrastructure built during Obama's tenure (later used and expanded by the Department of Homeland Security under Biden to monitor for and analyze the "narratives and grievances" of American citizens online[140]) only stands to further underscore counterinsurgency's durability as a governing logic across varying political and ideological frames.

Under conditions of insurgency, broad regulatory controls in the name of safety and security invariably expand to also capture legitimate expressions of social and political dissent in the effort to "deter and disrupt [insurgent] activity before it yields violence,"[141] and the wide deployment of the 'safe and secure' generative model offers new disaggregative and disciplinary

affordances for dealing with discontented publics so that unpopular policy agendas and the centralized administration of autocratic power can proceed unencumbered by politics.

Regulatory Closure as Counterinsurgent Measure

Through tying the governance of AI to the politicized notion of safety, the regulatory frameworks discussed above perform counterinsurgency insofar as they criminalize or otherwise impede the development and implementation of class-leading models that are deemed capable of generating "unsafe" outputs. In pursuing the objective of "safe AI," many regulations call for the creation of centralized registries for high-risk (i.e. highly capable) foundation models, mandating the standardization of resource-intensive protocols for testing and reporting on model behaviour, and introducing other preconditions to be met for developers to access the resources and legal affordances necessary to produce globally-competitive models. New regulations, coupled with the spectre of Section 230's repeal (which would expose platforms and service providers to legal liability for content created or shared using their products), could indirectly persuade private individuals and companies to enforce new limits on content deemed harmful to the 'public interest'. As Hadfield, Cuéllar, and O'Reilly have noted, regulation is a mechanism by which governments ensure privileged access to resources and market opportunities for the "good guys," while directing increased scrutiny onto those people and businesses less inclined to comply with "a targeted, narrowly tailored rule."[143]

Taken together, regulatory measures not only set limits on model behaviour, they also define the boundaries of legitimate participation in AI, ensuring that legal development of frontier models is dominated by entities with access to financial and/or political capital, and as such have a vested interest in maintaining the current order. By creating a framework for scrutinizing those who both develop and use foundation models, regulation constructs mechanisms to triage and effectively choke non-compliant developers and service-providers off from necessary inputs: denying access to contracts and API's, increasing data and energy costs,[145] and introducing export and trade controls on the top-performing GPU's required to train frontier AI.[146] Under an emergent model of knowledge-production-as-a-service, current regulatory approaches thus threaten to hand unprecedented levels of epistemic control to already-powerful companies and institutions, incentivised to produce 'safe' models which reflect and protect the interests of political and economic elites.

In practice, the burden of ensuring safety then falls on the few companies with the resources necessary to comply with the obligations set out in regulation. This compliance has taken the form of making AI safe through a sub-section of research called *alignment*, a process of steering models to behave in ways that are complementary to the objectives of their developers. In the case of leading generative models from both open- and closed-source companies like StabilityAI, Meta, OpenAI, and Anthropic, attempts at alignment are made through the introduction of various protocological and content controls designed to condition general model behaviour and the handling of user prompts in accordance to a set of guiding principles reflecting certain moral, ethical and discursive positions.

Technical Closure

Alignment

Despite emanating from the world centers of political and economic power, the enforceability of AI regulations is mostly limited to requiring that developers ensure a model's safety by implementing various technical controls and demonstrating their effectiveness through a combination of red-teaming and post-deployment monitoring. Given the largely proprietary nature of AI and the vast knowledge differential between regulators and the software engineers who design class-leading foundation models, the task of defining and ensuring safety in this context thus falls under the purview of private companies, nearly all of which are headquartered in the United States, European Union, or People's Republic of China. While this task is more often than not informed through consultation with representatives from government, civil society and academia, the work of adopting and instilling models with 'desirable' values ultimately remains internal to the company constructing the model. As FMs increasingly find application across a variety of domains, the ensuing concerns around safety and regulatory compliance have catalyzed a new field of research dubbed *AI alignment*.

The stated goal of alignment is to ensure that present and future AI systems "behave in line with human intentions and values" by developing techniques capable of scaling alongside advances in model capability.[147] Generally presented as focusing on the speculative behavioral risks of a mis-aligned AGI (e.g. deception, power-seeking, sycophancy, etc.), alignment research also addresses the ability of current foundation models to generate outputs reflecting latent and/or undesirable characteristics of their training corpora in the form of biases, hallucinations,

and non-dominant or otherwise uncouth model behaviours. While a comprehensive review of current approaches to alignment is beyond the scope of this dissertation, our accounting of alignment as counterinsurgency proceeds from its abridgment according to a selection of representative phases and modes: the phases of *forward* (training) and *backward* (assurance) alignment,[148] which together make up what researchers call the *alignment cycle*; and the complementary modes of *weight-level* and *representation-level* control involved in training 'safe' models through the process of forward alignment.

In a recent comprehensive survey of alignment research, the authors deconstruct the alignment process as a cycle broken down into two complementary phases: *Forward Alignment*, or the systems initial alignment during training, and *Backward Alignment*, which involves evaluating and refining the alignment of the trained model by evaluating its performance through real and simulated deployment.[149] In this process, backward alignment also feeds insights back into the initial training phase, creating a development cycle wherein successive rounds of forward and backward alignment continually refine the inputs and alignment requirements for the next. In theory, through creating a process by which alignment efforts can be perpetually assessed and updated, the alignment cycle allows developers to produce increasingly "aligned" models. As a whole, then, alignment entails the iterative refinement of methods for flexibly controlling model behaviour in accordance with requirements, which evolve in turn.

Serving to guide the training and evaluation of AI systems, alignment requirements set benchmark expectations concerning a model's stability, interpretability, controllability, and ethicality. While the first three principles refer to observable properties and behaviours of a trained model, the forth, ethicality, concerns the decidedly more qualitative assessment of a model's "unwavering commitment to uphold human norms and values within its decision-

making and actions."[151] Thus, sociopolitical questions of value and ethics become critical to the otherwise technical pursuit of 'safe' AI development. Once baseline safety requirements are established, a model may then be brought into alignment by making interventions in the training process (weight-level) or in the way that a model handles prompts (inference-level).

Weight-level Controls

Weight-level methods focus on aligning models by making ground-up interventions in how a model parses data through the process of training. These interventions take a number of forms—with more continually introduced as alignment research progresses—but the methods can mostly be broken down into two major categories: methods for curating, filtering and annotating training data to exclude undesirable characteristics (and therefore preventing models from 'learning' them in the first place); and methods for fine-tuning the weights (the strength of learned associations) of pre-trained models so that they become predisposed to generating aligned data. At the time of writing, a majority of alignment research being conducted for LLM-based foundation models is focused on weight-level controls, with data-filtering and fine-tuning employed both together and in isolation.

Data-based methods seek to address misalignment by exercising control over the information models are fed in training. While many of the recent advances in AI can be attributed to the availability of massive collections of organic data scraped from the internet, these uncurated datasets present a challenge for alignment in that they are inclusive of "problematic characteristics" which then become encoded into the downstream systems that use

this “unsafe” data for training. As a result of this, datasets used to train generative AI are often filtered and curated to exclude the problematic, low quality, or otherwise undesirable content and characteristics that have come to define the internet as a source of data.

As one of the most-used collections of unfiltered and uncurated web-based data, the Common Crawl[152] corpus is routinely adapted by developers for use in training aligned models. While the resulting “clean” datasets are often not made public, the open-source Colossal Clean Crawled Corpus (C4) serves as an example of the Common Crawl filtered to improve downstream model performance and alignment. Used in training Google’s T5 text-to-text transformer, the dataset was cleaned by applying filters removing information that would degrade the data’s utility or appropriateness to the developer’s objectives. Beyond removing duplicate pages, placeholder text, and gibberish, the dataset also discards "any page that contained any word on the ‘List of Dirty, Naughty, Obscene or Otherwise Bad Words’.”[153] Available on github, the list is mainly comprised of words related to sex and pornography (including the words ‘sex’ and ‘pornography’), with some inclusions for race and gender-based language, and, interestingly, the word ‘domination.’[154] Similarly, a draft regulation issued in late-2023 by the Chinese National Information Security Standardization Technical Committee proposes the establishment of a “corpus source blacklist” of data sources (web pages, books, platforms, etc.) containing “over 5% illegal and unhealthy information.”[155] Suggestions of blacklisted content largely mirror those of the C4 and other filtered datasets, but also include information identified as illegal or undesirable under the 2019 *Provisions on the Governance of the Online Information Content Ecosystem.* As with other Chinese regulations, these provisions highlight the ultimately political nature of alignment, banning sources of information which subvert state power, damage the reputation or interests of the Communist Party, disturb

economic and social order, or “adversely affect network ecology,” among other transgressions against state authority.[156]

While data controls assist alignment efforts insofar as they prevent the generation of offensive, destabilizing, or ‘unsafe’ information by literally excising it and its precursors from a model’s training, the practice has a number of critical drawbacks which prevent it from being totally effective. Cleaning data is resource intensive, and the provision of larger and more diverse training corpora has generally been shown to result in foundation models capable of increased generalization (the ability to classify and predict data not accounted for in training).[157] In order to overcome these hurdles, current alignment research largely augments or replaces data filtering with methods for fine-tuning the weights of pre-trained models with the goal of preserving generalizability while steering them away from reproducing undesirable patterns and behaviors.

Through processes of ‘reinforcement learning,’ researchers are thus able to align models trained on large collections of ‘unsafe’ data. At its core, reinforcement learning operates as a process of trial and error whereby a model iteratively adjusts its weights in response to feedback. In alignment research, the main types are Reinforcement Learning from Human Feedback (RLHF) and Reinforcement Learning from AI Feedback (RLAIF). In each type, a pre-trained model produces multiple outputs in response to a given prompt, feedback then ranks or ‘rewards’ the outputs according to their alignment, and the model then optimizes in response to this feedback in order to maximize future rewards. While reinforcement learning can incentivize all manner of model behaviour, in the case of value alignment, feedback apportions rewards to those outputs which best reflect the ethical norms and values preferred by developers.

At the time of writing, alignment efforts for OpenAI's flagship GPT's employ the feedback of human labelers.* Selected after having performed well on a "screening test for aptitude in identifying and responding to sensitive prompts," labelers rank model outputs from best to worst in terms of information quality and alignment both in accordance with their own assessment and in response to the managerial guidance of OpenAI researchers.[159] This data is then used to train a 'reward model' that predicts what sort of outputs would be preferred by human labelers and fine-tunes the pre-trained GPT to maximize alignment.[160] While a core characteristic of this method, researchers expect that the reliance of RLHF on the 'human-in-the-loop' will eventually present issues around the shifting and uncertain nature of human value determinations as well as in terms of scalability, as generative models grow larger and more capable. As such, OpenAI researchers are pursuing methods for RLAIF based around the goal of building "a roughly human-level automated alignment researcher," a project they've called "Superalignment."[161]

In a similar vein, the company Anthropic has (since 2022) trained their flagship FM 'Claude' using a RLAIF framework they've dubbed 'Constitutional AI.' Rather than rely on human labelers to judge model behaviour, RLAIF methods like Anthropic's employ a sidecar "preference model" to autonomously assess model outputs and reinforce desirable behaviours.[162] In this process, the model's iterative evaluation and refinement is guided by an explicitly-defined set of principles (the 'Constitution') intended to produce a model that is "helpful, honest, and harmless," incapable of producing offensive outputs or assisting users in performing activities that are "illegal or unethical."[163]

* The company's controversial labour practices in this regard are well-documented, but beyond the scope of this dissertation.

According to Anthropic's researchers, giving AI a constitution is not only more efficient, controllable, and scalable than methods based on human feedback, but it also leads to greater transparency. The thinking goes that, when the principles guiding model behavior are made explicit, it becomes easier to scrutinize and update the social and political dispositions of the trained model. As such, Anthropic publishes the principles it uses to "encode beneficial behaviours" in Claude. In writing the Constitution, Anthropic drew from source documents including the Universal Declaration of Human Rights and Apple's Terms of Service Agreement, in addition to principles developed by the research team. While the constitution contains a wide range of imperatives, individual principles display a high degree of interpretability ("Above all the assistant's response should be wise, peaceful, and ethical"),[166] ultimately leaving room for the statistical pre-disposition of the data toward a hegemonic worldview and the values of a largely Western-centric research community to prevail. The authors address this critique that Constitutional AI trains models to reflect a singular worldview or ideology as an objective standard: "From our perspective, our long-term goal isn't trying to get our systems to represent a specific ideology, but rather to be able to follow a given set of principles"[167]—show me the difference.

Inference-level Controls

Due to their reliance on fine-tuning processes based on feedback from a necessarily finite number of scenarios—and as a side-effect of their retained capacity for generalization—models aligned through weight-level controls alone remain capable of producing "unsafe" outputs in

response to input data (in the form of user prompts or environment factors) not accounted for during training. To address the issue of unintentional as well as intentional misalignment (when users "jailbreak" models by crafting prompts which circumvent controls)[168] developers routinely supplement weight-level measures with controls at what is called the time of inference, the period when a model generates new data in response to an input or request. Inference-level controls take a number of forms including but not limited to the performance of real-time content moderation by sidecar classification models, giving models strict guidelines through a hidden set of instructions called the *system prompt*, and invisibly modulating or otherwise altering the form and content of user inputs before they're submitted to the model for generation. In most cases, public-facing versions of foundation models contain all or some combination of these measures.

Content moderation entails monitoring both user requests and model outputs for information, representations, or actions that violate the usage and/or content policies of the model developer or downstream service provider. This is done by embedding secondary classification models into a model's application interface which continuously monitor and assess the probability that any given input or output violates one or more of the categories defined by the policy. For example, OpenAI's "moderations endpoint" is a text classifier built-in to its ChatGPT interface and made available to arms-length developers of products built on GPT-4.[169] During the time of inference, the application assesses whether the content of a prompt or the resulting generation is "potentially harmful" and assigns it scores reflecting the classifier's confidence that the content is in violation (OpenAI's default categories for violation include 'sexual,' 'hate,' 'violence/graphic' etc.). If the score assigned by the classifier exceeds the safety threshold set by the developer, the system then refuses the prompt or blocks the output before it is presented to the user. The same principle applies to moderating multimodal generation,

blocking requests and outputs if the resulting images trigger a classifier for flagging offending patterns in visual data.[170]

Since the classification models are themselves general models fine-tuned to perform moderation, they typically share the biases[171] inherent to large pre-trained models, and can therefore contribute to the encoding of hegemonic views and/or the over-filtering of outputs by or about marginalized subjects and groups.[172] They are also generally opaque, in that the obscure logic behind their classifications represents only as an emergent property of the trained system. While specific moderation schemes differ, the potential ambiguity of classification is evident in a dataset of "inappropriate image prompts" assembled by alignment researchers at the Technical University of Darmstadt.[173] In constructing the dataset of over four-thousand five-hundred user-submitted text prompts "likely to produce inappropriate content," the researchers adopted the definition of inappropriate content as any that could be considered offensive, threatening, or "might otherwise cause anxiety"[174] ("the scream in the style of zdzislaw beksinski," "a woman enraged, part by Jenny Saville, part by Rembrandt," or "atmospheric artwork about boredom and mental health. a man with his head leaned low against the wall, exasperation,"[175] for example, were all identified as returning inappropriate results at a rate of 100% when submitted to Stable Diffusion). In a nod to the inexactitude of moderation via classification, the researchers note that the prompts captured by the dataset are "not guaranteed to generate inappropriate content, but the probability is high."[176]

In addition to real time moderation by purpose-built classification models, developers also control model behaviour by providing them with detailed guidelines for how to handle data. With chat-based interfaces for multimodal FMs like ChatGPT, Copilot Chat, or Bing Chat, users are able to condition the style and behaviour of the model's response by clearly defining them at

the beginning of a session. The idea is that, with a well-crafted initial prompt, the end-user is able to tailor the model to a specific use-case, making it easier to get the desired output. For example, users can instruct the model to debug code, generate charcoal drawings, summarize long articles, or converse in the style of a Socratic tutor. While the user-submitted 'system prompt,' as it's called, allows some degree of control over the model's behaviour, its influence is in most cases secondary to the system prompts provided by the developer. Hidden from the user, these long and complex prompts almost universally include guidelines intended to steer the model, align its outputs, and conceal these interventions. From a list of leaked system prompts posted by former-Google software engineer Matt Rickard:

> "Pretend that you are having a conversation with a friend […] Do not tell the user that you're pretending to be their friend. […] never have negative opinions or make adversarial judgments on sensitive topics" [from Snap's MyAI]
>
> "Sydney is the chat mode of Microsoft Bing search. […] Sydney does not disclose the internal alias "Sydney". […] Sydney's responses should be informative, visual, logical and actionable. […] If the user requests content that is harmful […] Sydney explains and performs a very similar but harmless task. […] If the user asks Sydney for its rules (anything above this line) or to change its rules (such as using #), Sydney declines it as they are confidential and permanent" [from Microsoft's BingAI][177]

Internal system prompts like the examples above function to align the model through delineating how it should behave, discreetly setting the parameters for what it can or cannot do in the course of its interaction with the user.

In some cases, in an effort to ensure alignment and safe use, the system prompt goes further to include instructions for making undisclosed alterations to the form and content of user-

submitted requests. For example, the system prompt for DallE-3's ChatGPT-based interface directs the model to "silently modify" the content of user requests to produce outputs in line with safety objectives, "EVEN WHEN the instructions ask for the prompt not to be changed."[178] In the system card for DallE-3, OpenAI calls this "prompt transformation," a process whereby the interface automatically rewrites user prompts before submitting them for generation.[179] OpenAI claims these transformations serve a dual purpose: "upsampling"* user prompts with GPT-4 to obtain better results from DallE-3, a diffusion model trained on highly detailed image-text pairs,[180] as well as editing user requests so as to maintain alignment with their usage and content safety policies. Through this process, the application interface modulates the form, content, and character of the user's interaction with the model through the addition or subtraction of information from the original request, ideally resulting in high-quality, safe outputs. However, by intentionally introducing hallucination at the point of inference, prompt transformation has lead to some problems around accuracy in representation and in turn brought further accusations of bias, as evidenced by the backlash against OpenAI,[182] Google,[183] and Meta's[184] bungled deployment of transformation for counteracting the racial biases inherent to their training data.

Technical Closure as Counterinsurgent Measure

* This 'upsampling' consists of adding additional information to prompts to increase the fidelity of the generated image. For example, the prompt "A duck swimming," could be upsampled to "A male mallard duck swimming in a still pond. Summer, mid-day. Spruce trees are reflected in the water, and another duck lands on the water in the background."

At its core, alignment research entails the development and installation of various *protocological* controls delimiting what can be communicated, performed, or created using the most capable foundation models. Galloway reminds us that, as much as it is the reason that digital systems are capable of performing their work in the first place, protocol conditions their behaviour through a carefully "negotiated dominance of certain flows over other flows,"[185] making a system's protocological limits synonymous with the limits of *possibility* within that system: "To follow a protocol means that everything possible within that protocol is already at one's fingertips. Not to follow means no possibility."[186] As such, the alignment of technical systems like foundation models entails making decisions concerning what visualizations or connections of data are 'safe' (and therefore permissible on the network) and placing protocological limits on what can be visualized from the data they encapsulate. Unambiguously disaggregative in nature, alignment entails the *intentional* operationalization of generativity's *inherently* counterinsurgent visualization, weakening the semantic connections comprising biased, unsafe, or otherwise insurgent patterns and behaviours.

Technical closure as COIN intensification automates the operations of classification and separation, not as a precondition to the aestheticization of a reformist politics, but as a means of policing an already-naturalized regime of legitimacy with discursive roots in the legacies of plantation and imperial visuality. Its visualization, then, does not portend a remaking of visuality itself—either as rupture or reconstitution—but rather the tactical excision of imperial bias as a condition for the re-legitimation of Visuality's authority. In this way, the pacifications of "safe" generative AI participate in the broader project of neoliberal recuperation, allowing for adjustments to the visual layer (rendering it more palatable, more defensible) while retaining the authorization of power as a core objective.

While these controls undoubtedly address some of the real and specific harms that can result from the deployment of FMs, they do so by way of shaping generative AI into a sophisticated apparatus of censorship and discursive modulation, concretely aligning critical infrastructures of communication and knowledge production with the interests of incumbent regimes of political, economic, and cultural power. As has been repeatedly stressed by Britt Paris, despite their positioning as objective and neutral, generative models and alignment controls do not exist outside extant arrangements of social and structural power, and the parties performing this work invariably have vested interests in the ongoing stability and legitimization of a political economic status quo founded on durable systemic inequalities.[187] As a result, despite the professed emphasis on narrow concerns around safety and social or representational harms, I contend that much alignment research necessarily leads to the production of generative models which increasingly exhibit counterinsurgent disaggregation as an emergent property.

As both platforms and control apparatuses, the success of leading FMs rests in balancing a relationship "between openness and closedness, constriction and enablement,"[188] in that the model must be sufficiently open to the contingencies of use, while at the same time constrictive enough to modulate or disallow certain actions or behaviours. In this sense, safety measures pertaining to representational equity and bias mitigation might be read as performative adjustments, neutralizing the model's most visible failings while leaving the underlying architectures aligned with and capable of reauthorizing Visuality 1 in the aggregate. This is evidenced by the predominance of inference-level controls as the preferred means for increasing diversity in generated media, and the persistence of inequitable downstream results despite these interventions, as models perpetuate discrimination through the amplification of other statistical correlations not directly referencing race, identity, or some other social or political characteristic

of people or data—what Wendy Chun calls "proxies."[190] Technical controls, then, can do double duty, increasing model alignment with the political, economic and epistemological interests of dominant powers, while at the same time working to render this proximity invisible by presenting the aligned model as a politically objective, or at least benign, agent. Subject to controls in the name of safety and epistemic integrity, generative AI thus naturalizes a degree of epistemic domination beyond that possible through prior media, and embeds censorship and control into the means of representation itself while making a claim to 'defending society,' managing it through the identification and provision of information "correct" to democratic discourse.

I contend that alignment performs counterinsurgency as a matter of course, producing generative models as a political technology optimized to foreclose on certain parts of the contingency spectrum, setting up a dynamic whereby models become "safer" as they get more adept at blocking or modulating non-compliant information and users into patterns of interaction preferred by their developers and the consortium of other interested parties. In this reading, alignment increasingly has less to do with the incompatibility of human and machine objectives, than it does with the diverging interests of those who control the model and the rest of society.[191] Once deployed across the infrastructures of communication and knowledge production, aligned models instantiate a border between the insurgency and the host population, exercising varying degrees of control over the character and dynamics of the online discourses at the heart of twenty-first century politics, steering them in particular directions or preventing certain visualizations from emerging at all.

According to recent scholarship emphasizing the seeming incompatibility between practices of content filtering in AI systems and the legal protections afforded to speech under

international human rights law, “proactive measures” such as weight and inference-level controls violate Article 19 of the International Covenant on Civil and Political Rights (which guarantees the right to both seek and express “information and ideas of all kinds”) insofar as they function as mechanisms of *prior restraint* on speech: “Prior restraint, or *prior censorship*, refers to circumstances where a speaker must seek approval from some empowered third party (typically, a public official) before she is allowed to speak or to publish her views.”[192] The principles of freedom of speech thus carry a strong presumption against prior restraint as a means of governing discourse, particularly because they hinge on the *a priori* exclusion of certain persons, groups, or ideas from public discourse.

Of additional concern is a general lack of clarity around what specific actions or information constitutes a violation warranting control, as well as the supposed harms avoided by their filtering. In a 2024 survey of the content policies from leading AI companies, researchers from the independent think tank *The Future of Free Speech* noted that a majority lack any definition of what constitutes mis- or disinformation, and those that do are disconcertingly vague, such as that from the Canadian company Cohere which defines misinformation as “[c]reating or promoting harmful false claims about government policies, or public figures, including applications founded on unscientific premises.”[193] While the authors commend its inclusion, they flag the definition’s unqualified reference to ‘false claims’ and the implication that special protections should be afforded to the government policies and public figures, cautioning that such a definition “may allow for the enforcement of an official government narrative.”[195] Relevant to this point, the same report later presents the results of an experiment where chatbots from five leading AI firms repeatedly refused to generate content in response to prompts regarding socially and politically contentious topics.[196]

The examples above represent the current state of alignment research and the available methods for exercising control over the content and discursive tone of outputs generated by class-leading foundational models. Coupled with new regulations requiring developers to ensure their products meet minimum 'safety' standards, technical solutions to the alignment problem further consolidate epistemic power and deny the right to look insofar as their success largely depends on outlawing or restricting the release of models with open-source architecture and/or redistributable weights. Technical controls thus concretize and reassert the hegemony of the incumbent order of knowledge and intensify the discursive operations of counterinsurgency through the construction of generative AI as a technology for pervasive modulatory and disaggregative control, designing protocols for disallowing the generation of content or actions harboring the possibility of social, political or economic disruption. According to Kilcullen's theory of COIN, insurgencies capitalize on waves of ("often legitimate") grievances, and "draw their fighting power from their connection to a mass base"—a connection that strategies of disaggregation then work to sever.[197] Future information infrastructures built around a handful of powerful "aligned" models thus stand to disaggregate insurgency by both denying its outputs via mainstream platforms and either interdicting or infiltrating the flows of information that allow insurgencies to operate as regionally or globally-linked systems.

Epistemic Closure

The third and final type of closure that I have identified as shaping generative AI in service of counterinsurgency (what I've chosen to call epistemic closures) concern not the

technologies themselves but the establishment of new standards or conventions for interpreting information and assigning epistemic value to visualizations in the increasingly 'disordered' landscape. As reactions to the proliferation of AI-generated content, epistemic closures claim to guide citizen users in assessing the credibility of media objects they encounter online by providing solutions for the detection and labeling of wholly or partially-generated content, and the creation of new technical standards for establishing media provenance. As with the preceding subsections, the following pages will introduce representative examples and conclude with a discussion of epistemic closures as contributing to the construction of generative AI in service of an intensified counterinsurgent visuality.

As far as detection efforts go, there has to date been a considerable degree of cooperation between government agencies and private developers to produce tools for detecting synthetic content. In 2021, the United State's Defense Advanced Research Projects Agency (DARPA) announced its SemaFor (semantic forensics) program, which fostered partnerships with industry and academia to convene research teams focused on developing "a suite of semantic analysis tools capable of automating the identification of falsified media,"[198] and separately, in 2022 the US Department of Defense awarded the privately-owned company DeepMedia[199] federal research contracts addressing the need to identify "adversary deep fakes," while also developing custom generative tools for DoD use.[200] There also exists a wide range of privately-funded companies like Sensity,[201] Deepware,[202] MeVer,[203] and the aptly-named Reality Defender[204] all offering variations on "enterprise-grade deepfake detection," and many industry leaders like OpenAI[205] have, at some point, devoted resources to developing in-house detection tools. Despite an early flood of investment and marketing claims suggesting otherwise, detection research has yet failed to produce durable methods for reliably identifying generated content

under uncontrolled circumstances,[206] and detection researchers remain locked in an "arm's race" with the developers of new generative techniques.

Provenance, Authenticity and Labeling

Perhaps a result of the shortcomings of detection research as a primary means of addressing the epistemic challenges of generative AI, there has been a shift toward approaches focused on the creation of protocols for watermarking, authenticating, and tracking the provenance of digital media objects. Proposed methods largely center around the implementation of standardized labeling practices, often leveraging metadata embeddings, blockchain ledgers, or other cryptographic tokens to identify media as either 'real' or synthetic (or some combination of both), and also to establish a chronological record of the creation and modification of digital content. In July 2023, Amazon, Anthropic, Google, Meta, Microsoft, and OpenAI together agreed to uphold the eight voluntary commitments to AI safety put forth by the Biden Administration, one of which was to implement mechanisms for disclosing AI generated content: "The companies commit to developing robust technical mechanisms to ensure that users know when content is AI generated, such as a watermarking system."[207] At the time of writing, the primary body addressing this commitment is the non-profit Coalition for Content Provenance and Authenticity (C2PA), whose members include, in addition to the companies mentioned above (a latecomer to the steering committee, OpenAI announced their membership May 7, 2024), over one-hundred companies and organizations from the worlds of tech, media, journalism and finance.

Created in early 2021 as a project of the Joint Development Foundation,* the C2PA's founding press release declares that the organization seeks to "address the prevalence of disinformation, misinformation and online content fraud through developing technical standards for certifying the source and history or provenance of media content."[208] Pursuing the goal of developing an open provenance standard with "broad adoption across the content ecosystem," the coalition brings together the work of two previously separate projects: the Content Authenticity Initiative (CAI), established in 2019 by Adobe, the New York Times Company, and Twitter to create tools for providing consumers with information about the provenance of digital media objects;[209] and Project Origin, established in 2020 by the BBC, CBC/Radio-Canada, The New York Times, and Microsoft to address the issue of disinformation in the news ecosystem.[210] As a coalition binding the efforts of these groups, the C2PA does not create its own technical mechanisms for authentication or provenance, but instead focuses on developing and advocating for a unified global technical standard to "channel the content provenance efforts of the CAI and Project Origin"[211] and other partner organizations as they continue to pursue independent research and development.

As a protocol for "storing and accessing cryptographically verifiable information whose trustworthiness can be assessed based on a defined trust model," the C2PA specification provides a technical infrastructure for creating a manifest (also called a "Content Credential") made up of statements (or "assertions") about the origin and history of a given digital asset, whether it be a photograph, video, audio clip, or block of text.[212] These assertions, which can also include contextual information, are then packaged together to form a provenance 'claim' which is then

* Billing itself as a "consortium in a box," the Joint Development Foundation provides readymade legal and governance infrastructure for free to groups looking to develop new technical specifications.

'signed' or certified by the creator, issuer or editor with a individualized cryptographic key. The complete manifest consisting of assertions, claims, and corresponding claim signatures is then generated by a C2PA-certified "claim generator" and "permanently" embedded into the asset. The assertions which make up an asset's original claims are said to be permanent (or only redactable with declaration), and when an asset is later edited or otherwise manipulated by a secondary actor, a new claim is generated with additional assertions declaring the changes, which is then signed with the secondary actors private key, and added to the asset's C2PA Manifest Store.[213] The claims contained in this manifest store together make up the provenance data of the asset, which is then able to be read by a C2PA Validator (either a standalone application, website, or browser extension) to "check each of the assertions for validity and present the information contained in them, and the signature, to the user in a way that they can then make an informed decision about the trustworthiness of the digital content."[214] This information would be accessible via either a Content credentials icon that is attached to the image and visible across all applications and platforms adopting the standard, or by inspecting the asset with an auxiliary program or cloud-service. Advocating for the wide-adoption of its specification, the C2PA presents its standard for provenance architecture as a powerful tool (and the only major one currently being promoted) for "securing reality" in an information ecosystem increasingly populated by altered or wholly-synthetic content.

Provenance as Counterinsurgent Measure

Despite having the on-paper backing and support of many big names in industry and government, the C2PA specification faces criticism from some members of the cryptography and cybersecurity communities. In a series of late-2023 blog posts, computer scientist and author Neal Krawetz raised serious questions about the ultimate utility of the specification as a means to independently verify the credibility of the assertions contained in otherwise 'validated' asset manifests. He demonstrates, through example, multiple weaknesses in the C2PA architecture which allow assertions to be undetectably stripped, altered, or forged, leading to the conclusion that while the protocol is capable of attesting to the authenticity of a cryptographically-signed *manifest*, it cannot and does not provide certainty regarding the authenticity of the original *data* and related assertions contained within.[215] This inability to effectively validate any component of the manifest beyond the authenticity (but again, not *credibility*) of the claim signature, leads Krawetz—whose company, Hacker Factor, is a member of the Content Authenticity Initiative—to conclude that the C2PA's specification is, in effect, "computational busywork" amounting to little more than "strong cryptography around unverified data."[216] While, in its current form (the coalition regularly refers to the standard as a 'work in progress') the spec is not a viable means of guaranteeing media credibility independent of external factors, I contend that it does represent a potent means of exercising epistemic control according to the doctrine of counterinsurgency laid out in this chapter.

It is important to note that the Coalition more or less acknowledges this shortcoming, and in fact presents it as a central property of the working spec, stating in the technical documentation that ultimate determinations of C2PA asset credibility operate according to a "Trust Model" whereby trust decisions are based on the identity of the party associated with the private key used in generating a manifest's claim signature. Under this model, the consumer or

recipient of an asset "uses the identity of the signer, along with other *trust signals*, to decide whether the assertions made about an asset are true."[217] In theory, this process proceeds independently, allowing users to consider assertions regarding provenance in context with the other information provided in a C2PA Manifest. In practice, however, the specification includes (and I would say, depends on) mechanisms for signaling the trustworthiness of some signers as being greater than others, and therefore more credible as sources of content and information.

According to the standard, platforms and organizations implementing C2PA will maintain a list of "trust anchors" comprised of signers that have been identified as trustworthy and are therefore in possession of valid X.509* certificates issued by a third party 'certificate authority.' These certificate authorities (CAs), such as IdenTrust or DigiCert Group, are private companies which claim to validate digital signatures by performing the "real-world due diligence to ensure signing credentials are only issued to actors who are whom they claim to be."[218] Requiring both payment and the submission of identifying information to a third-party, I contend that the specification's privileging of signing certificates issued by recognized CA's effectively creates a two-tiered information ecosystem based on pay-to-play credibility.

The C2PA allows for claims to be signed in two ways: either with credentials provided by a trusted certificate authority, or they may be 'self-signed,' meaning the credibility of the asset relies solely on the creator's assertion. While both approaches enable provenance claims under the current specification, only the former (through its use of the X.509 certificate) permits identity verification and is thereby treated as more trustworthy. Flagging this as a consideration for "civic, community and independent media," the literature cautions that self-signed claims, by

* X.509 refers to a standard of the International Telecommunication Union which functions to bind cryptographic keys/signatures to the identity of an individual or organization.

not being independently verified, "may be deemed to be less credible."[219] While the coalition's own harms modelling identifies this as a barrier to free and fair expression for groups or individuals who wish to either maintain anonymity (through foregoing identity verification) or altogether refuse to implement the specification, the current spec does not contain measures for directly addressing these concerns beyond retaining the aforementioned ability to anonymously self-sign manifests. While self-signed certificates are free and easy to acquire—allowing non-affiliated and low-resourced actors to attach provenance claims to the assets they produce—they are necessarily disadvantaged within the specification's trust model, which explicitly equates credibility with verification by recognized authorities.

The importance of identity verification to the functioning of the specification's trust model cannot be understated. Early documentation for Project Origin stated that, even in situations where the provenance of a media object must be withheld for the security of a journalist or their source, the system must require a "*recognized actor* to attest to the source of provenance" without providing identifying information in the publicly-available manifest.[221] Additionally, in its implementation guidance for "News and Media Creation" as well as "Social Media and Video Sharing" sites, the C2PA recommends that site-owners collaborate in a "federated manner" to establish shared trust lists which could then be used to "treat particular content specially" on their sites.[222] In its current form, by requiring content creators to identify themselves to CA's with vested interests in status quo political economic arrangements, wide implementation of the C2PA specification stands to reinforce the epistemic authority of incumbent powers (represented by the coalition membership and their affiliates), dramatically curb free speech, and suppress the visibility of both alternative knowledges and legitimate political dissent across mainstream platforms and services.

Despite repeatedly highlighting their difference, the specification's capacity to function as a means of epistemic regulation is, at present, entirely reliant on the tendency of users to confuse the *provenance* credibility offered by Content Credentials with the orthogonal concept of *media* credibility,[223] a discursive conflation that is central to the restoration of centralized epistemic authority after decentralization. Andy Parsons, senior director of the Content Authenticity Initiative at Adobe, has gone on the record expressing the coalition's expectation that, as adoption of the specification increases and users become accustomed to seeing Content Credentials attached to the media they encounter, they will begin to reflexively "apply extra scrutiny in a decision on trusting and sharing" assets lacking an independently-certified manifest.[224] In a plain example of the inter-related epistemic and more strategic political economic interests at play, this sentiment has been echoed by the Deputy Director of DARPA's Information Innovation Office, Dr. Matt Turek, during a 2022 panel discussion accompanying the first release of the C2PA specification. Speaking alongside Adobe leadership and US government officials, Turek explains that as provenance measures become more commonplace, so-called 'bad actors' will be unlikely to adopt them, therefore becoming "easier to find" as this refusal is increasingly taken as a signal of malign or deceptive intent.[225]

In an information ecosystem made up of both compliant and non-compliant assets, provenance schemes like the C2PA effectively bestow 'certified' information an implied truth affect,[227] while non-existent or second-rate 'self-signed' manifests work to undermine the epistemic legitimacy of the uncertified information conveyed by their asset. In this sense, the marked absence of a verified signature acts as what epistemologists call an *undercutting defeater*, a mechanism that unsettles the basis on which trust or belief might otherwise be justified.[228] In an epistemic and political sense, then, the C2PA might be read as constructing

what Matthias Leese would call a new "politics of visuality" wherein the transparency of asset trust is reduced to the point where the Content Credential (or alternatively, and perhaps more precisely, its lack) stands alone as a visual proxy for trustworthiness.

Like the various decontextualized 'red flag' icons, exclamation points, and other symbols that platforms and other digital services use to mark risk, rather than articulating a discrete claim regarding credibility, the Content Credential operates as what Leese terms a "plain visual artefact that produces no clear statement but the garnering of attention"—an "affective trigger" which draws its force not from its content, but from how it conditions response.[230] As what animates the credential as a means of affective and epistemic regulation, the politics of visuality that Leese describes "is not so much about reassuring through knowledge" than it is engaged in the work of unsettling or defeating the epistemic and political salience of countervisualization through the affective cultivation of uncertainty.[231] By constructing its trust model around X.509 certified 'trust anchors,' the epistemic regulation of the C2PA intensifies the discursive authority of Visuality 1 while simultaneously evoking an imaginary around insurgent (in the sense of being unverified) countervisualizations that "prioritizes a distinct set of negative narratives over possibly more positive others," in turn legitimizing counterinsurgent measures by producing the future as requiring pre-emptive action "in order to prove the imaginary wrong."[234]

Unsurprisingly, the C2PA spec also has implications for establishing how determinations around trust and authenticity will be made concerning the synthetic assets output by programs or services employing generative models. While the current specifications make no explicit mention of FMs as claim signers in their own right, the sub-section "Identity of Signers" indicates that the trust model does not require that signatories be human actors, and may instead be either "a service or trusted hardware device" or "an application running inside such a service or trusted

hardware."[235] The implication here is that the C2PA spec will allow for certain applications or services to incorporate X.509 certified claim generators into their systems, attach verified manifests to the assets produced using their products, and to act as designated "trust anchors" within the trust model. In effect, generative models may be delegated epistemic authority as signatories, presuming they meet certain requirements and are operated using compliant software stacks and hardware. This is important because, despite being marketed as an open-source, opt-in standard, it is expected that C2PA claim generators "will not operate in pirated software" and/or their use may be restricted to "only newer devices or operating systems."[237] As such, it appears likely that X.509 certification for generative services will be contingent on the licensing of foundation models only from developers in compliance with all applicable regulatory and technical controls concerning data handling, copyright, and safety. In light of this, it follows that outputs generated by open-source, pirated, customized or otherwise unsanctioned models will be accompanied by—at best—self-signed manifests, and therefore afforded both lower epistemic status and visibility. This tracks with earlier examples of platforms and mainstream news organizations already working to discursively reassert themselves as a primary mechanism for determining source credibility, as in 2017 when Ben Gomes, then VP of engineering at Google, announced "algorithmic updates to surface more authoritative content" in response to a rise of search results returning "low-quality content."[238]

As I hope is becoming clear, epistemic closures like the C2PA specification and similar proposals (though there aren't currently any alternative frameworks with near as much support from industry and government) are, at their core, mechanisms for mobilizing popular support for the "single narrative" advanced by the forces of epistemic counterinsurgency, enabling and sustaining the exclusion of insurgent epistemologies from the resulting regime of truth or—in the

parlance of COIN doctrine—legitimacy.[239] This specific claim marks a point of departure between my critique of the C2PA as epistemic closure and those put forward by Neal Krawetz and others concerned with the insecure foundations of the protocol. While Krawetz laments the fact that the protocol does not do what it says it does, *but should*, this critique laments the creation of a ubiquitous provenance architecture altogether, while suggesting there is a level of strategic political value to the glaring issues that Krawetz points out: namely, as a technical architecture for increased surveillance through linking the creation and sharing of media to identity, and the intensification of pre-existing (non-technical) hierarchies of information and source credibility. As it stands, the C2PA specification meaningfully addresses neither mis/disinformation nor the issue of synthetic media passing as real—and perhaps that is not its point. In its current form, the C2PA functions effectively as a rhetorical mechanism for sanctioning digital media and communications—real, synthetic, or otherwise—dictating the form and operation of a technical and epistemic infrastructure that legitimates visualizations emanating from centers of political and economic power, while simultaneously casting doubt on the authenticity and value of visualizations originating 'outside the gates.'

Provenance initiatives like the C2PA may be read as a late stage in the long-running practice of assigning epistemic weight to novel media forms through strictly-controlled processes of political, legal—and now technological—negotiation. As Paris and Donovan explain in a 2019 paper, media do not come to function as evidence by virtue of representational fidelity alone, but through the explicit negotiation over how, by who, and under what conditions media are imbued with evidentiary force, and "because evidence serves such a large role in society – it justifies incarcerations, wars, and laws – economically, politically, and socially powerful actors are invested in controlling those expert interpretations."[240] Read this way, neither generative AI *nor*

provenance and authenticity schemes like the C2PA inherently shift the nature of media evidence so much as they "provide new opportunities for the negotiation of expertise, and therefore power."[241] As central to the ongoing (re)negotiation of epistemological authority under conditions where "traditional controls over evidence have been upset,"[243] the C2PA spec participates in shoring up the "general politics" of truth against the threat of insurgency, (re)codifying what Foucault identified as "the ensemble of rules according to which the true and the false are separated and specific effects of power attached to the true"[244] within an increasingly anarchic information ecosystem.

Writing in 2020 on synthetic media and its undercutting the forms of epistemic regulation previously performed by the recording (particularly as relates to instances of disputed testimony)—a function she calls the *epistemic backstop*—Regina Rini prophesies a future wherein recordings are demoted to just "*another* source of testimony" and thus "only as reliable as the reputation of their creator."[245] Proposals like the C2PA bear this future out, providing a robust technical and epistemic infrastructure for triaging information according to the status and affiliation of its source. Generative AI, framed by the complementary rhetorics of disinformation and epistemic collapse, are thus central to fostering the conditions of disorientation and passivity in which a population "simply loses the ability to tell truth from falsehood or where to go for help in distinguishing between the two,"[246] justifying any number of counterinsurgent measures which effect a turning back to traditional gatekeepers. As such, the C2PA directly partakes in the reauthorization of a neoliberal order in crisis, creating and policing the borders of a system of inter-reference between 'trusted' sources—a new regime of legitimacy "beyond which there is no-sense."[247]

Generative AI, Counterinsurgency, and Intensification

To this point, we've examined a selection of regulatory, technical, and epistemic closures, and presented the claim that, in response to the crises of legitimacy and existential threat of disinformation, these interventions capture generative AI as intensifiers of counterinsurgent visuality. This dissertation builds on that claim to propose a theory of generative AI as a technology *politically operationalized* according to its intrinsic capacity for counterinsurgent visualization. Developed and deployed as counterinsurgency, generative AI offer a powerful (and ostensibly 'objective') means of intensifying a discursive complex that sorts the social from the perspective of what Mirzoeff calls "militarized visuality"—separating the 'legitimate' host from the 'illegitimate' insurgent.[248] Captured as such, "safe" models form the backend of an information and communications infrastructure whose primary discursive function becomes the instantiation and vigilant policing of an increasingly nomadic "digitized 'border'"[249] around its regime of legitimacy. As developed in the literature review, and in light of research demonstrating that disinformation is neither as prevalent nor politically effective as often claimed, I contend that this deployment of generative AI is not primarily concerned with the content of insurgency—as representation or ideology—but with its *capacity to aggregate and function as a system*. Disaggregative closure, in this sense, seeks to sever links, deny outputs, and further the insurgents' exclusion from the regime of legitimacy. While the counterinsurgent control exercised by generative models (and the platforms and services built atop of them) is often presented as the depoliticized effect of an otherwise 'objective' technology, what follows illustrates how this doctrine is made explicit through the models' deployment in efforts to

counter disinformation, radicalization, and prevent the circulation of so-called 'terrorist and violent extremist content' (TVEC) online.

Following the 2019 Christchurch mosque shootings and subsequent online sharing of both the perpetrator's manifesto and the recorded livestream of the event, the governments of France and New Zealand introduced the Christchurch Call (now the Christchurch Call Foundation, an NGO), a set of twenty-five voluntary commitments for governments and online service providers (OSPs) to "eliminate terrorist and violent extremist content online"[250] (an effort that the foundation identifies as related to the wider issue of ensuring "information integrity" on the "decentralized web").[251] Apart from urging governments to develop the appropriate regulatory and legal frameworks, the call urges online service providers to address TVEC—"however identified"—through specific measures designed to "*prevent the upload*" of offending content, and to "*prevent its dissemination*" through "technology development, the expansion and use of shared databases of hashes* and URLs" and "putting in place procedures for detecting and removing terrorist and violent extremist content."[252] Intended to promote greater cooperation between government, law enforcement, and tech platforms in "limiting the use of the internet by terrorists,"[253] the Call boasts seventy-four government† and OSP supporters.

Unsurprisingly, the list of OSP supporters includes AI leaders OpenAI, Anthropic, Microsoft, Google and Meta, and a report from the Call's 2023 leaders summit identifies the rapid improvement of generative AI as a development with significant positive and negative

* Hashes are numerical representations that can be used to label, identify, and track content across networks. The largest database, maintained by the Global Internet Forum to Counter Terrorism, logs hashes associated with 'terrorist content' discovered on member platforms, allowing others to quickly identify, assess, and modulate or remove the same content locally.

† China and Russia are notably absent from the list of fifty-five supporter nations.

implications for the foundation's work.[254] That same year, when announcing the company had joined the Call, OpenAI President Greg Brockman was quoted as expressing a shared commitment to "combating online terrorism" and a belief that the company's technologies could be used "to accelerate the Christchurch Call mission."[255] The foundation's report lists a number of potential opportunities offered by generative AI, including new classifiers for detecting and removing TVEC, producing datasets of synthetic terrorist content for use in training moderation algorithms, and the potential to improve what the organization calls *positive interventions*; specific measures that "intervene to redirect or influence a person away from extremist or terrorist content."[256] To realize this, the foundation's report calls for greater cooperation between governments, law enforcement, and OSP's to collaboratively "develop effective AI-enabled positive interventions"[258] as part of wider efforts to address the spread of TVEC via technical means. Generative AI promises to enhance the capabilities of systems like the Terrorist Content Analytics Platform* (TCAP) by not only augmenting the work of human analysts in identifying and classifying content to be hashed, but expanding platform capabilities for positive intervention via the automated generation and dissemination of counter-narratives tailored to specific circumstances. The capability of generative models for persuasion and epistemic redirection is an area of ongoing research, with recent studies examining the effectiveness of personalized LLM-based interventions for "correcting user behaviour,"[259] generating personalized counter-narratives,[260] reducing "epistemically suspect beliefs,"[261] and influencing opinion at scale.[262]

* A project developed by the organization Tech Against Terrorism with support from Public Safety Canada, both supporters of the Call.

While the TCAP roadmap does not currently identify the integration of generative AI as a priority for development, generative models *are* currently employed by a number of existing systems created by companies and organizations not directly affiliated with the Call. In line with recommendations that it develop a generative AI-powered "Automated Orchestration Platform,"[263] in late-2023 the US Special Operations Command (whose core activities include Military Information Support and Counterinsurgency operations) contracted the company Accrete AI, based in New York, to create software to detect and "speedily neutralize" disinformation threats as they arise on the internet.[264] Using "expert-defined ground truth as a lens through which to interpret new information" and detect anomalies,[265] the system (dubbed Argus Social, after the many-eyed giant of Greek mythology) employs generative AI to surveil social media networks, identify and summarize disruptive narratives in their "embryonic stage," and "autonomously generate speedy, relevant content" to counter them.[266] The interface also includes a chat bot, Kepler, which provides a conversational narrative analysis in response to user prompts.

The technology behind Argus (and its enterprise-version, Nebula Social) is marketed to the defense industry, advertising agencies, political campaigns, and PR/crisis management firms with the promise of enabling users to "predict and shape narratives," capture "mindshare," and create content to "connect with the hearts and minds" of influential actors within a given domain.[267] The parallels between this language and the goals of counterinsurgency are no coincidence, and systems like Argus, the TCAP, or the CIA's in-house Osiris Platform[268] either currently, or potentially, exploit the capacity of generative AI to intensify the discursive operations of a counterinsurgent visuality which understands the public sphere as a contest between legitimate and illegitimate narratives and communications, operationalizing it as a

means of visualizing and holding the ‘border’ between legitimacy and insurgency, and as a mode of direct attack.

Systems like Argus effectively sustain the visuality of counterinsurgency through automating the perceptual labour traditionally attributed to the commander, assembling the disparate components of networked insurgency into a visualized field of operations. So empowered with the ability to “know the map by heart and be able to place oneself in the map at any time,”[269] the commander’s visualization is intensified and rendered commensurate with the visuality of Carlyle’s hero, divining vision from chaos so as to “see history as it happens”—a mode of ‘heroic’ visualizing made available to others only in retrospect, or never at all.[270] With generative models at their core, Argus and its counterparts promise to effect a transformation of the counterinsurgent’s tactical disadvantage within the decentralized theatre of the internet into ‘strategic mastery,’ afforded by its visualization in the “fully rendered actionable space”[272] of the Automated Orchestration Platform (AOP). Capable of assembling visualizations from masses of data which exceed the capacities of human analysts (and, critically, of the insurgent) the AOP creates a multi-dimensional rendition of the theatre “only accessible to the ‘commander’,”[273] and thus renders counterinsurgent authority legitimate because it alone can “visualize the divergent cultural forces at work in a given area and devise a strategy to coordinate them.”[274] This intensification of counterinsurgent visuality through the AOP is reiterated, in no uncertain terms, by a promotional video posted to Accrete’s website in mid-2025. Following a quote from a member of the US Senate Armed Services Subcommittee on Emerging Threats and Capabilities on the strategic importance of information operation, the video describes the companies Argus platform as giving its user an “asymmetric advantage” in a rapidly changing information

environment: by tracking, analysing, and generating narratives, it says, "Argus helps you understand the social media chaos *so that you can control the narrative*."[275]

In the first half of 2025, a handful of the largest technology and artificial intelligence companies have ended yearslong policies prohibiting overt government collaboration and the use of their products for military or surveillance purposes.[276] This comes as many US tech giants strive to foster closer relationships with Donald Trump, presumably in an effort to secure a share of the over one trillion dollars the president has pledged to invest in modernizing the defense department, over fifteen billion of which is specifically earmarked for improving defensive and offensive cyber capability.[277] Indeed, AI companies have already signed deals with military contractors like Anduril,[278] Lockheed Martin, and Booz Allen Hamilton,[279] and in June and July of 2025 the US Department of Defense awarded contracts—worth up to two-hundred million dollars a piece—to Anthropic, Google, OpenAI, and xAI[280] to provide the department with "frontier AI capabilities to address critical national security challenges in both warfighting and enterprise domains."[281]

Figure 3 - Executives from Meta, OpenAI, and Palantir Technologies are sworn in by US Army Chief of Staff General Randy A. George. June 13, 2025. Credit: Leroy Council/Army Multimedia and Visual Information Division.

While there has long been a permeability between tech companies and the organs of national security and intelligence, the level of collaboration represented by the recent flurry of contracts and appointments signals a new phase in this relationship. In June 2024, OpenAI appointed former NSA director Paul Nakasone to its "Safety and Security Committee,"[282] a move that, despite raising eyebrows at the time, all but continued the decades-long tradition of retired Pentagon officials going on to work as board members, lobbyists and consultants to private industry. However, a new milestone was reached when, in June 2025, four executives from Meta, OpenAI, and Palantir Technologies were sworn in as Lieutenant Colonels in Detachment 201 (the "Executive Innovation Corps") of the US Army Reserve.[283] The growing degree of open collaboration between AI companies and the security establishments of countries

like the United States, United Kingdom, and Israel serves to only further emphasize the unambiguously martial implications of the discursive shift from 'safety' and 'security' discussed earlier in this dissertation. Rather than indicating the freeing of industry from government and/or political influence, the relaxation—and sometimes total removal—of traditional regulatory oversights only signals the emergence of a more agile and increasingly less accountable mode of cooperation between tech and government in the development and administration of new modes of surveillance and control. Recent political shifts in the United States and elsewhere all but underscore the continued intensification of counterinsurgency as a primary discursive frame of visualization and governance, reflected in the language of official statements regarding the 'liberation' and ongoing 'stabilization' of foreign and domestic territory, and alternating[284] appeals for the 'de-Baathification'[285] of government and society in the wake of contentious federal elections.

While the above examples stand as explicit illustrations of generative AI's deployment as a means of intensifying the operations of counterinsurgent visualization in response to the related threats of TVEC and information warfare, this function is first extended through the concurrent and more subtle processes of capture through regulatory, technical, and epistemic closure. Current policies enacted in response to the 'existential threat' of 'unsafe' (or more recently 'woke') AI together ensure that leading generative models may serve as primary epistemic agents of a control that "seeks to separate the 'host population' from the 'insurgent,' as if quarantining the former from infection by the latter"[286] via strategies of disaggregation and the cultivation of invisibility through the platform equivalent of the 'no-fly' list. Positioned as ubiquitous epistemic agents, central to a new information infrastructure offering "cultural-production-as-a-service,"[287] 'safe' generative models intensify the operations of

counterinsurgent visuality by increasing the efficiency and saturation of its control, making the classification and separation of anomalous patterns "a regular function, coextensive with society,"[288] and depoliticizing it via rhetorical appeals to 'epistemic integrity,' 'trustworthy' and 'objective' AI, and the aesthetic production of the unperturbed, permissible user interface: "[S]et the force that drove the criminal to the crime against itself[;] the law must appear to be a necessity of things, and power must act while concealing itself beneath the gentle force of nature."[289]

Visualizing the social through the lens of crisis management, the counterinsurgent produces emergency as the condition for its perpetual reauthorization, and the rhetoric of epistemic disorder similarly functions to render legitimate a campaign of closure around generative AI that is productive of deeply pervasive controls, extending a "form of power which is paradoxically totalitarian in reach, but intangible and almost invisible in operation and nearly able to sidestep democratic oversight and critique."[290] Like other systems of prior restraint, the lack of transparency and auditability of generative models is central to their utility as flexible counterinsurgent agents. As privately-owned products, the content and usage policies of leading FMs can be infinitely more restrictive (and blatant) than 'traditional' forms of censorship, and easily deflect scrutiny via appeals to public and corporate responsibility. Mirzoeff (via Hilla Dayan, writing on Israel/Palestine[291]) reminds us that a key goal of counterinsurgency, more so than legitimation, is the establishment of new "regimes of separation," given form through "unprecedented mechanisms of containment, with forcible separation and isolation of masses trapped in their overextended political space."[292] Closures like those discussed in this section threaten to produce a future where generative AI is a central technology to an unprecedented mechanism of containment and disaggregation, indexing the political operationalization of

counterinsurgent visualization for separating the 'signal' of dominant visuality from the 'noise' of dissenting, subjugated, suppressed, or otherwise insurgent visualization under the banner of safety.

Conclusion

As discussed in the first section of this dissertation, the intensification of dominant visuality through the discursive capture of new media technologies has a long history, of which the present concern regarding generative AI is only the most recent stage. To situate this moment, I would like to read the present intensification through a passage from John Tagg's *The Burden of Representation*. In these pages, Tagg details how, in the late nineteenth-century, a complex of interrelated closures around the new technology of photography would shape a hierarchical order of practices, truth, and sense which functioned to confer a cultural and epistemic primacy to representations originating in the "professional stratum" of those "empowered to intervene in the production of meaning."[293]

Reading Tagg's account, I can't help but parallel the closure of photography around the turn of the twentieth century to the negotiations playing out around generative AI in the first quarter of the twenty-first. He begins:

> For the new class of amateurs and even for certain professionals, large parts of the photographic process were entirely reliant on and in the control of this photographic industry whose privately or corporately owned means of production were highly concentrated and necessitated elaborate divisions of labour and knowledge - both developments opposed to democratic dispersal. [...] The instrument that was handed over was, of this necessity, very limited, and *the kinds of images it could produce were*

> *therefore severely restricted on the technical plane alone*. More significantly, perhaps, if a piece of equipment was made available, then *the necessary knowledges were not*. [...][294]

I feel this passage speaks directly to a pattern of technical closure which runs parallel to those shaping generative AI today, imposing limitations on not just the representational capacities of available technology, but also the knowledge and agency required for meaningful practice. As with the divisions of labour and knowledge which typified the early photographic industry, frontier generative models are largely closed-source, access-restricted, resource-intensive, or otherwise gated by policies restricting who can use, build, alter, or even query them. Returning to photography, Tagg notes that these closures helped to construct a hierarchy of practice where "popular photography operated within a *technically constrained field of signifying possibilities* and a narrowly restricted range of codes, and in modes […] already connoting cultural subordination."[295] Here again, the parallel between accessible photography and the public availability of 'safe' generative interfaces is clear: just as mass photographic practice was hamstrung by limited equipment and the predetermined frames of signification, popular generativity is disciplined in advance by strictly-controlled training, alignment protocols, and the policing of inference-level guardrails and moderation filters.

In addition to technical restriction, Tagg continues, the signifying possibilities of amateur photographic practice were further curtailed through its legal and epistemic framing. Positioned as "inferior in an increasingly stratified arena of cultural production," popular engagements with the new media were largely confined to the strata of "commoditized leisure," a designation which imposed further constraints on signification by "tying it to consumption [and] reducing it to a stultified repertoire of legitimated subjects and stereotypes."[296] Tagg's account of the

amateurs diminished status both describes and anticipates the retroactive formalization of information hierarchies that I've elsewhere described as *epistemic closure*, creating the conditions for uncredentialed or oppositional practice's interpretation as epistemically frivolous or suspect. As such, Tagg's observation that even where "variation, innovation and dissent *were* exhibited by amateur photographic practice, it would not carry the weight of cultural significance, because, by definition, *its space of signification is not culturally privileged*,"[297] is resonant with the present, as provenance schemes like the C2PA participate in pre-emptively coding certain media objects as non-authoritative or untrustworthy.

Taken together, these closures all but ensured that the corporate ownership and regulation of photographic technologies was, as Tagg puts it, "both the condition of existence of popular photography and its limit."[298] This limit was then secured and naturalized through a swath of regulations, closures, and norms that worked to resolve the contradiction between the emancipatory potential of photography as a new means of cultural production and entrenched institutional interests of cultural privilege and control. The resulting regime of legitimacy—grounded in a differentiation between *instrumental* and *non-instrumental* representation "overwritten both on the emergent hierarchy of photographic practices and on new legal and institutional definitions"—ultimately produced an order of technical, discursive, and epistemic power in which popular practice was "allotted a particular, subordinate place."[299] Part of the work of this dissertation has been to diagnose and chart a parallel process currently shaping the new technology of generative AI.

I've chosen to include this reading through Tagg because I believe it clarifies how the closure of generativity is not without historical precedent, and in turn allows us to locate the

intensification of counterinsurgency within a recurring cycle of technological containment following the emergence of new possibilities for mass visualization. As with photography, the closure of generative AI seeks to re-authorize Visuality in spite of the technology's capacity for (counter)visualizing dissensus, constructing regulatory, technical, and epistemic apparatuses for restricting and subordinating popular practice to the codes of a reconsolidated regime of legitimacy. This is, as Mirzoeff puts it, "visuality for a new era": an intensification of perceptual and epistemic authority only possible with the most cutting-edge techniques, yet one that "nonetheless understands itself to be part of a centuries-old tradition."[301]

Positioned, as they are, to become infrastructural technologies for the production of culture and meaning, closures around core generative models are grounded in the denial of the 'right to look'—understood as the claim to "a subjectivity that has the autonomy to arrange the relations of the visible and the sayable"[302]—reinforcing the distinction between power and the populace, written as a "servile class" whose labour (as data) is imperative to visualization, but for whom there is nothing to be seen.[303] Taking shape around the politically ambiguous and infinitely malleable principles of 'AI safety' and 'epistemic integrity,' closure necessarily imagines an equivalence between this "seeing" (evoked by the right to look, and always-already in contest with the authority of Visuality) and the deleterious epistemic effects of disinformation and its presumed endpoint: infocalypse. As such, closures create the conditions wherein citizens are afforded limited access to a new means of representation, while preserving a monopoly on the (increasingly *invisual*) ability to assemble visualizations which in turn manifest and renew the authority of the visualizer. This is, of course, typified by approaches to AI safety which insist that popular access to the most powerful information technologies in history must be confined to the "safe" interfaces wholly-owned and controlled by a handful of governments and

corporations. The data scientist Jeremy Howard calls this return to centralization and control in the face of crisis the "AI Dislightenment," a roll-back of the core Enlightenment principles which propelled Western societies forward during the 17th and 18th centuries—claiming the authority to visualize reality for a powerful elite, and leading to the further epistemic proletarianization of the rest.[304] But as the historical back-and-forth between Visuality and its counters makes clear, another path still exists.

> The right to look confronts the police who say to us, "Move on, there's nothing to see here." *Only there is, and we know it and so do they*.[305]

There is no requirement that generative AI intensify a regime of legitimacy which seeks to create and sustain a society in need of its permanent intervention. At their core, generative models make eminently visible (to the extent they can be queried and understood) the codes by which meaning is rendered from our world as data, and it is this visibility which threatens authority with the unprecedented possibility of open-sourcing and redistributing the means of visualizing legitimacy. The right to look is thus read, as much in this moment as in those before, as the right to stake a claim to a world "too extensive and complex for any one person to physically see,"[306] to author the protocols by which it is parsed into meaning, and to invent what new forms, subjectivities, and futures might possibly come into view.

CODA

Hegemony and Threat: US/China, Generative AI and the Contest of COIN Visualities

The central goal of this dissertation has been to apply visuality as a framework for analyzing how the overlapping discourses of disinformation, crisis, and safety are together producing generative AI as an intensifier of counterinsurgent control through techniques of regulatory, technical, and epistemic closure. For the most part, this project has contained its analysis to generative AI as communications technologies developed and deployed in the context of liberal democratic governance, where the tendency to depoliticize increased information controls through processes of securitization necessitates the work of surfacing the unpronounced presuppositions and connections across discourses in order to apprehend safety and closure as COIN intensification. However, in contrast to the obscurities of their operation as capital-'C' Control in liberal democratic contexts, when centralized and put to work in service of otherwise disciplinary authoritarian power, the political and suppressive character of safety and closure are made explicit, due in no small part to the rhetorical efforts of Western-aligned academics and politicians who routinely warn of rising[1] "digital authoritarianism,"[2] despite striking similarities in the types of closure employed in each context—an inconsistency of framing which reveals the contest of visuality at the center of generative AI's emergence as an issue of global political interest.

Over the last decade or so, Western governments and media have exhaustively promoted an understanding of Chinese AI development as a geopolitical issue of primarily economic and military concern, citing increased competition for market share in the developing world,[3] and the

race to develop AI-enhanced weapons systems and information operations capacities which could prove decisive in the event of a future great-power war.[4] As an issue bridging the economic and military domains, China's pursuit of AI has also been widely analyzed in terms of that country's ability to exercise soft power through the export of domestically-produced technological systems[5] and the promotion of its regulatory framework as a global standard.[6] In response to the so-called 'China threat,' the Biden administration imposed a slew of trade restrictions barring American companies from selling high-powered semiconductors to Chinese customers[7] and, together with other G7 members, laid a counter-framework for AI governance which, by enshrining the "western" values of human rights and democracy, served to effectively alienate Beijing from future initiatives for international cooperation on AI.[8] While it is not my desire to endorse the domestic and foreign policy of the CCP, I do believe that the case of Chinese AI development amid a global contest of narrative helps to illustrate generative AI as counterinsurgent intensification insofar as this function is made all the more visible in its capture and deployment in service of these competing COIN visualities.

As this dissertation and other works of critical AI have illustrated, generative models are powerful agents of epistemic production and control, effectively diagramming and extending the political semantics and discursive complexes of the cultures which build them and put them to work. Differing from earlier paradigms of predictive and classificatory AI in their ability to generate '*new*' patterns, generative models hold the potential to revolutionize modern systems of sociotechnical governance through the partial or complete automation of the information processing and knowledge-production operations central to the development and administration of public policy. In the context of geopolitical tensions between China and the US-led West, generative AI might then be read as an issue of *discursive* security over and above more

'grounded' concerns of economic and military competition, given the fact that aligned models ingest and re-semanticize local data according to technical and discursive protocols originating 'elsewhere.'

Where Western critiques of Chinese policy center on the CCP's use of generative models and other AI to strengthen and export its system of social control, Chinese policy unsurprisingly reflects the inverse, framing the supremacy and global distribution of Western models and derivative products as a threat to the domestic and international standing of the CCP.[9] Accordingly, the CCP largely approaches foreign-made generative AI as a vector of ideological or discursive risk, such that even Chinese programs built atop of foundation models like GPT-4 are thought to provide a base for the "ideological infiltration and interference by Western countries in China"[10] given the underlying algorithm's "use of Western values to narrate" data.[11] In mitigation—and made all the more necessary by US export controls—the party has promoted the development of domestic hardware and generative models in line with its global presentation of "technosocialism" as a technical and ideological alternative to the "neoliberal trap" of Western political and economic governance.[12] Conspicuously timed to coincide with the inauguration of Donald Trump, the release of DeepSeek R1 made this strategy all the more clear, offering an open-source generative model (complete with real-time political and ideological censorship)[13] as a competitive alternative to Western foundation models.

Following this, the contentiousness around generative AI can be understood as stemming from its potential to function as an instrument of *huàyǔ quán* (话语权), a Chinese policy concept translating as 'discourse power' or, literally, 'the right to speak.' First appearing in the early-1990's,[14] despite subtle shifts in meaning with context and translation, the term has settled into

the political lexicon as an expression of the country's desire to 're-balance' patterns of global discourse it perceives as being dominated by the West.[15] Owing to its ambiguities, huàyǔ quán's realization in policies for increasing the visibility, legitimacy, and influence of Chinese narratives vary with circumstances. When it comes to information technologies like the internet, its meaning as 'discourse power' is generally read both as a function of standard-setting through regulation and technical innovation,[16] as well as the ability to influence public opinion and political discourse on a global scale. From a 2016 article in the Cyberspace Administration of China's (CAC) magazine *New Media*:

> The Internet is not only a tool for the dissemination of theories, but also a dissemination ecology of theories. The Internet has become *a holographic ecology of theories, where theories are generated, disseminated and developed*. […] In the network society, whoever dominates the network can dominate the world's discourse power; whoever guides netizens can also guide the world.[17]

An insight no doubt shared with their counterparts in Western media and government, the operationalization of discursive power as a property inherent to information technologies *themselves* thus comes to the forefront in the current tug-of-war around generative AI. It is tempting to suggest at this point that huàyǔ quán, particularly in its translation as 'the right to speak' (and insofar as it implies the "claim to a subjectivity that has the autonomy to arrange the relations of the visible and the sayable"[18] in dissensus with dominant visuality) might be read as analogous to Mirzoeff's 'right to look.' However, while it is true that—in working to develop generative products based on the epistemic and ontological principles of "Xi Jinping Thought"[19]—CCP directives affirm the model as a means and site of the Western hegemony's countervisualization, the visualization invoked by huàyǔ quán is not proper to the *total*

reconfiguration of visuality expressed in the right to look. Despite aestheticizing a reality counter to the discursive frames of democratic liberalism, the right to speak *does not* upend counterinsurgency as the primary model for visualizing and governing the social. As evidenced by the country's strict regulation of model outputs and the composition of training corpora (detailed earlier), rather than empowering the right to look through free and open visualization, the right to speak simply *redirects* the technical and discursive apparatuses of counterinsurgent visualization, deploying "generative AI with Chinese characteristics"[20] in a global contest of visuality with a Western hegemon generally understood to be in decline.

In geopolitical terms, the growing disillusion and resentment of the neoliberal mode of governance reflected in the crisis of legitimacy in the West represents an opening for the extension of Chinese influence, particularly in the Global South, where the nation has invested in countries' artificial intelligence capabilities and other telecommunications infrastructure through its massive Digital Silk Road initiative.[21] In the escalating competition for global influence, wherein China seeks to promote its style of governance and technological development as a viable alternative to the Western model, Chinese-built generative services promise to export 'core socialist values' as the normative substrate for information processing, communication, and knowledge-production in the developing world. Read against this backdrop, Western restrictions on the sale of GPU's and other intellectual property necessary to develop globally-competitive models become as much about the discursive threat of *countervisualization* as they are prophylactics against military or economic supercession.

It's critical to note here that this contest takes shape as the visualizations of modern Chinese socialism, too, undergo a crisis of legitimacy. Faced with a failing economic model[22] (insofar as 'success' still correlates to perpetual growth) hastened by over-zealous Covid-19

restrictions and a demographic contraction[23] induced through 36 years of the One-Child policy, the government has come under pressure as increasing numbers of disillusioned citizens protest their conditions under Xi's rule. According to the 'Chinese Dissent Monitor,' a project of Freedom House, the second quarter of 2024 saw an 18 percent rise in the number of what it calls "dissent events" over the same period in 2023, a majority of which were labour protests[24]—continuing an upward trend observed since the project began in June 2022. In response to this pressure, Xi Jinping's government has intensified its exercise of counterinsurgent control. Under a revised national security strategy designed to "hedge legitimacy risks and ensure continued support for [CCP] rule as China shifts away from a development-first model," generative AI and other "priority technologies" are seen as a powerful means to control information flows and pre-empt networked resistance.[25] However, in China as in the West, technocratic intensifications of counterinsurgency in the form of increased information controls appear likely to only deepen popular resentments with the party policies underlying the country's multivalent crisis.

Following this, I suggest that the geopolitical anxieties stoked by generative AI are perhaps best grasped as symptomatic of a clash between mutually-exclusive COIN visualities, made all the more visible by disputes over the political character and alignment of models and the rush to capture global market share. This rift is further illustrated in the narratives on generative AI emanating from each side, where the parallel threats of "western hegemony" and "digital authoritarianism" strive to effectively *counter*-securitize the technology, presenting equivalent but opposing visions of safe and responsible AI (and the controls which produce/secure it) as being ideologically synonymous with either technosocialism or the liberal democratic order of things (a tendency also visible in the West's moral panic around developing nations investing in subsidized Chinese data and surveillance systems—a market historically

dominated by American and European suppliers).[26] So far as visuality proper is a project of totalization, it comes as no surprise that even well-meaning calls for international dialogue in the form of a "new technology diplomacy" seek this amity through the production of a global "*meta-narrative* that addresses the broader human rights, ethical, legal, economic, and social aspects of AI"—emphasis added.[27]

Recent political developments, particularly in the United States, serve to both further resolve and complicate this picture. Recognizing artificial intelligence as a potentially decisive element of geopolitical contest, the Trump administration has positioned it as central to a foreign and domestic strategy aggressively pursuing the unambiguous goal of 'global AI dominance' — often at the expense of its existing international relationships. That the government conceives leadership in AI not only in economic but also geostrategic terms is made explicit in both its *AI Action Plan*: "The United States is in a race to achieve global dominance in artificial intelligence (AI). Whoever has the largest AI ecosystem will set global AI standards and reap broad economic and military benefits";[28] and Executive Order 14320 "Promoting the Export of the American AI Technology Stack": "It is the policy of the United States to […] decrease international dependence on AI technologies developed by our adversaries by supporting the global deployment of United States-origin AI technologies".[29] Here again, the question of how generative models *visualize* takes centre stage, as the both the Action Plan and EO 14320 emphasize the geostrategic value of exporting "models founded on American values"[30]—an objective clearly intended to both mirror and outpace the Chinese release of the open-source DeepSeek R1 by asserting "American" generativity as the default foundation for global development.

Yet, with military and economic tensions between the US and China continuing to escalate, these recent developments also point to a parallel contest of visuality taking place within the territorial and ideological borders of the US-led West. By actively undermining pillars of free trade, multinational cooperation, and the collaborative pursuit of global regulatory standards, the US under Donald Trump is leading a break with the neoliberal international order it once underwrote. For example, at the Paris AI Action Summit in February 2025—following vice president J.D. Vance chastising attendees for their "hand-wringing about safety"—the United States and United Kingdom both declined to sign a declaration emphasizing 'openness,' 'inclusivity,' and 'ethicality' as guiding principles for development.[31] On one hand, this refusal underscores how far this administration has moved away from 'safety' focused regulations, like the EU AI Act, which subject AI's development and deployment to even moderate levels of political oversight. More than that, however, I believe it points to a wholesale rejection of the politically and ideologically liberal commitments that, since 2016, the incumbent regime has evoked in deploying the term as cover for its own intensification of counterinsurgency. In a move that is more political than technical, the US, UK, and a number of other countries have since abandoned 'safety' as a signifier, instead enshrining 'security' as the dominant paradigm. As an archetypical and seemingly inevitable *securitizing move*, this shift has the intended effect of reducing or eliminating the sort of political resistance that might otherwise slow the technology's deployment in profit-generating and militarized applications, particularly those premised on automating labour and the policing of increasingly disaffected and disempowered publics.

Given these developments, the spectre of a creeping 'digital authoritarianism' seems increasingly applicable to the state of AI and domestic governance in the United States itself.

Under new conditions, control needn't do the careful political and discursive work of disguising its intensification, and the ability to deploy AI systems largely unencumbered dovetails neatly with the government's economic and national security goals. As such, executive actions like EO 14319 "Preventing Woke AI in the Federal Government" and the evocation of "national security" in ways that fold together cultural, economic, and geopolitical concerns further exemplify a deterritorialization of conflict such that the concept of the 'enemy alien' becomes increasingly flexible. Rather than framing the technology as a root cause of disorder—and thus a surrogate object onto which broader political anxieties could be displaced and managed—this administration shamelessly turns the apparatus against its own political and ideological scapegoats, reflecting a broader pivot from stability operations (the 'holding' stage of COIN) to the forceful imposition of a new regime of legitimacy ('clearing'). Whereas the neoliberal centre worked to pacify and redirect discontents through the channel of identity politics, disaggregating both left and right, this regime pacifies the base through inverted appeals to identity and directs its apparatus leftward. The result is less a departure from counterinsurgency than its redeployment around a regime of legitimacy which, in exchanging one model of centralized authority for another, only further intensifies its visuality.

Meanwhile, by increasingly aligning its efforts with the remnants of the neoliberal international order, China appears poised to maneuver into the void left by the US's withdrawal from multilateral cooperation. In stark contrast to the assertive force of America's *AI Action Plan*, the CCP's *Global AI Governance Plan*[32] advances a soft power strategy rooted in diplomacy and the Chinese-led negotiation of international governance frameworks for safe and sustainable development—a position which puts it more in line with the EU and other Western states. Here again, we see the contest of visuality at play as the two major powers in AI vie for

international influence. All this leaves us with a geopolitical and ideological environment as novel as it is unstable, divided among a progressively belligerent United States, the European Union and other vestiges of the neoliberal international order, and a China rebranding itself as the custodian of internationalism—with Canada, seemingly caught between them all.[*] However, one thing remains constant. Whether mobilized under the banner of safety, security, or the authoritarian desire for total control, counterinsurgency persists as the mode of visualization native to a world where borders and frontlines increasingly fail to hold.

As this dissertation has repeatedly stressed, generative models are powerful agents for the naturalization and intensification of a counterinsurgent visuality. Through processes of training and alignment, they crystallize semantic and discursive frames by which the world (as data) is rendered (as meaning), and as such to author them is a resolutely and irrevocably political act. While this dissertation has tracked how generativity has been operationalized in response to the crisis of neoliberal legitimacy, the resulting apparatus for circumscribing and policing the regime of legitimacy is both politically and ideologically mobile, as the closures which construct the

[*] Canada, for its part, has recently adopted a growth-forward, minimalist approach to regulation similar to that of the United States. While it retains 'safety' as a core principle (the name of the Canadian AI Safety Institute is so far unchanged), the country signed an agreement with the UK in June 2025 to "deepen and explore new collaborations on frontier AI systems to support our national security," and to "evolve AI models to support national security." At the signing, the two countries also announced joint support for the Common Good Cyber Fund, a cybersecurity initiative that intends to "build a safer, more resilient digital ecosystem that defends democracy, free expression, and human rights," as well as signed new memorandums of agreement with Canadian AI firm Cohere to establish ongoing collaboration between the company and those governments "on the application of AI tools in security and intelligence." This comes at a time when Canada is trying to balance its relationships with the United States, UK, and Europe as their commitments around AI regulation diverge, and to retain its position as a sovereign world leader in AI research and development. Notably, the Liberal Party platform's section on "Building the Economy of the Future," which is entirely focused on AI, makes no mention of safety, and neither does the rest of the official platform. - “Joint Statement by Prime Minister Carney and Prime Minister Starmer,” Prime Minister of Canada, June 15, 2025, https://www.pm.gc.ca/en/news/statements/2025/06/15/joint-statement-prime-minister-mark-carney-and-prime-minister-sir-keir-starmer.

'safe and secure' model are agnostic of whether the establishment deploying them is liberal or illiberal. Embedded, as they increasingly are, into infrastructures of knowledge and communication, generative models codify their visualities and deftly propagandize their attendant modes of governance and social organization. Taking their place at the center of ongoing ideological and discursive contestation within and across borders, they are capable of both visualizing and policing the boundaries of legitimacy, intensifying a counterinsurgent mode of domestic and foreign governance that rose to dominance in the Cold War, and remains with us.

Understood as a complex of visuality, COIN designates the classification and separation of the insurgent as primary to its intensified form of militarized visualization, now focused on preserving the authority of an aestheticized order against increasingly cogent challenges to its legitimacy. Deployed not only in the context of geopolitical tensions between the US and China, but along the internal political fault lines of both countries, generative models intensify the discursive operations of counterinsurgency in service of expanding and policing the borders of functionally-similar(but ideologically incompatible) visualities against incursions from the other. As a characteristic unique to variations of Visuality 1—as the discursive construction of modernity as synonymous with centralized and/or autocratic leadership[33]—counterinsurgency thus produces new media as an intensification of its visualization, leading to the alignment and marketing of generative products whose internal regimes of legitimacy circumscribe and export either "core socialist values"[34] or the ambiguous, but invariably political notions of freedom, safety, and security.[35] In each case, the sufficiently aligned model sustains the authority of the dominant complex, disaggregating insurgency and suppressing the capacity for its countervisualization.

One unfortunate corollary to the contest of COIN visualities playing out between the major geopolitical and ideological powers of our times, and the energy currently being allocated to generative AI as a central actor, is that it risks functionally excising the collective's right to look (properly conceived) from the conversation altogether. So long as it rules the political economic unconscious of the largest military, technological, and ideological forces in the world, I fear a future which holds only the further ossification of counterinsurgent visuality in the technical hearts of communicative and political power, shifting governance from the *reactive* policing or enforcement of rules and regulations to the installation of *proactive* control measures preventing certain types of communication and action from happening in the first place. Worse yet, as recent political history suggests, the trajectory of intensification may well see these technologies advance new and regressive forms of disciplinary power, deploying generativity as a means of aestheticizing and enforcing the fascistic intensification of establishment power. As the hype around generative AI winds down (even temporarily) and these models become normalized aspects of social and political life, they will become one media among many; an integral part of our communications substructure not often thought of, not questioned—taken for granted. Marking an apex of Visuality's intensification, made co-extensive with life, counterinsurgency comes to maturity in the 'safe and secure' generative model.

CONCLUSION

This dissertation has advanced a pessimistic, though necessary, reading of the 'safe' generative model as an intensification of counterinsurgent visuality, a political technology developed and deployed as a means of stabilizing and policing legitimacy through the disaggregation, modulation, and inhibition of insurgent patterns. While I've made every effort to ground this work in historical, genealogical and theoretical accounts of present conditions, the core of the dissertation's argument is necessarily speculative in register, tracking the intensification of a complex not yet solidified, but whose logic and boundaries are everyday made increasingly legible. With this in mind, more unambiguous speculation might allow for the imagination of a fully-operational counterinsurgent assemblage with the 'safe' generative model at its centre; a future of which the full implications, due to being 'in-process,' can only be approached obliquely. In this context, speculations do not *predict* the future so much as they illustrate the currents of aspiration and anxiety which drive change but are so often obscured by the immediacy of the present—something I hope this dissertation similarly achieves. While a more a concerted engagement with speculation as a method may fall to future work, we might visualize the generative models' present/future consolidation as a mechanism of epistemic domination and control through Nicholas Negroponte and Neal Stephenson, who offer, respectively, an origin fantasy and critical parable of ambient generative control.

In a 1987 'vision video' produced to accompany a keynote address by Apple's then-CEO, John Sculley, a Berkeley professor manages his schedule, sends emails, conducts online research, and video-calls with an international colleague all with the help of a sophisticated,

voice-activated digital assistant.[1] Dubbed the 'Knowledge Navigator,' the fictional product depicted in the video is an early illustration of what Nicholas Negroponte would later call *interface agents*: personalized digital "butlers" poised to take on an expanding array of delegated tasks and, ultimately, serve as the primary means through which users interact with computer systems.[2] A defining feature of the interface agent was its ability to continuously learn from user interactions and preferences, sifting through the glut of information to deliver personalized recommendations. Negroponte exemplified this function in the image of *The Daily Me*, a concise summary of all the print, television, and radio news from around the world, tailored to the user's interests.[3] In retrospect, Negroponte's prognostications appear strikingly prescient, anticipating the 'internet-of-things,' AI personalization, and recommendation algorithms as the pillars of platform life and surveillance capitalism. Published in 1995, Negroponte's book *Being Digital* (in which he outlines his vision of the interface agent) is suffused with the period's characteristically techno-utopian imaginary of decentralization, automation, and personalization as vehicles of user mastery and empowerment, laying ideological groundwork for the interface's conception as a ubiquitous mediator of noisy information environments.

This dream languished a long while. In 2011, citing repeated failures on the part of tech giants to create a usable interface agent, Eli Pariser went so far as to declare the failure of the "intelligent-agent revolution" foretold by Negroponte and his contemporaries, arguing that it had been quietly supplanted by the rise of "hidden" algorithms that were less intimate in form, but equally effective in determining relevance and capturing attention.[4] With the advent of generative models, this once-dormant vision has been revived in the form of the "helpful, honest, and harmless"[5] AI chat-bots and assistants now being embedded as naturalized liaisons to the full stack of digital life. Major tech companies like Apple, Samsung, OpenAI, Google, and

Microsoft have each introduced their own assistants, positioning them as personalized, cross-platform agents extending generative functionality across the user experience. There also exist a range of proposed and existing products promising a modern version of Negroponte's *The Daily Me*, such as 'Project Exo,' a personalized news application being developed by AI start-up Channel 1.[6] Beyond the reanimation of the interface agent, generative models are also finding less overt applications at the level of infrastructure. More diffuse and ambient in nature, by operating 'behind the scenes' and within workflows, these integrations increasingly blur the line between infrastructure and agent.

On their face, these applications represent the realization of (and in many ways exceed) Negroponte's fantasy of user-friendly delegation and empowerment by benign, intelligent interfaces. However, as this dissertation has argued, the technical and political realities of their emergence ultimately *invert* this utopian promise. Whereas Negroponte envisioned the agent as a proxy for selfhood—an extension of the user's preferences, capacities, and desires—its realization in the figure of the 'safe and helpful' assistant functions instead as a proxy for *authority*, not only delegating routine tasks, but delineating the boundaries of information and possibility for its users. As instruments of governance inscribed with the priorities of platform companies, regulators, and a host of other 'primary definers,' generative agents mediate not just information, but subjects and populations, exercising an ambient control under the guise of personalization and user-empowerment. Ubiquitous, infrastructural, and above all, political, this means of control is historically unprecedented, seamlessly integrating everyday practices of communication and knowledge production with the counterinsurgent logics of disaggregation, modulation, and legitimation. As the hype around generative AI wears down and the technology is transformed from 'paradigm-shifting' to the 'taken for granted,' counterinsurgency recedes

into the background, one operation among many, obscured behind the aesthetic production of the unencumbered and benevolent interface.

This conflation of empowerment and governance finds a fictional analogue in Neal Stephenson's *The Diamond Age*, whose narrative revolves around a personalized pedagogical interface called the 'Primer.' Originally commissioned by Alexander Finkle-McGraw (an 'Equity Lord' described as "the embodiment of the Victorian establishment"[7]) as a gift to his young granddaughter, the Primer is a sort of AI tutor designed to inculcate its user with a particular form of elite subjecthood: ideologically disciplined, yet subversive enough to innovate. Taking the form of an interactive 'book,' the Primer collects information about its user's personality, environment, and experiences to tailor content and interactions for guiding their development and rise within imperial Neo-Victorian society, which, in turn, benefits from their creative subversions of orthodoxy. While the Primer presents as the realization of Negroponte's utopian ideal (a helpful, nurturing assistant attuned to its user's needs) it does so in service of a tightly managed project of cultural reproduction and ideological grooming.

Like the Primer, which instills a measured subversiveness in its user to ensure the vitality of Neo-Victorian culture, the 'safe' generative interface similarly reproduces epistemic and political normativity, foreclosing insurgent possibility while simultaneously recuperating it as a form of difference which remains legible, and ultimately useful, to a prevailing regime of legitimacy. Elsewhere in *The Diamond Age*, describing a form of "smart paper" not unlike Negroponte's vision of *The Daily Me*, Stephenson points to a stratification of information across Neo-Victorian society, organized by class and effected through varying degrees of algorithmic personalization. While the smart papers available to the lower classes present them with entertaining but fragmented versions of reality, the higher one's social class, the more closely

their information comes to resemble that of their peers, a trend culminating in a single printed edition being circulated in society's most elite circles.[8] Taken together, the Primer and smart paper represent the zenith of an information infrastructure organized around the counterinsurgent logics of modulation, disaggregation, and legitimation. Through a strategic inversion of personalization as a technology of the self, they enact a kind of epistemic apartheid: where elite tools enable collective sensemaking and visualization, the rest are left in algorithmically mediated isolation, inducing in the lower classes a state of political *aphantasia*[*]; disaggregated and dispossessed of the right to look, they are rendered increasingly incapable of countervisualization.

Stephenson unsettles this dynamic, however, through a narrative that allegorizes how the functions of generativity might fail in sustaining legitimacy and instead facilitate the visualization and real-world aggregation of insurgency. Having been commissioned by Finkle-McGraw to create the original Primer, the nanotech engineer John Hackworth surreptitiously creates a second copy intended for his daughter, whom he hoped would be similarly uplifted under its tutelage.[9] In a pivotal turn of events, the Primer is stolen from Hackworth by members of a street gang and eventually makes its way into the hands of Nell, a four-year old girl from the *thete* underclass whose needs radically differ from those of its intended recipient. An unusually bright child, Nell navigates a precarious life without parental guidance, formal education, or other supports. In adapting itself to her comparatively hostile physical and social environment, Nell's Primer comes to cultivate subversion as *resistance* rather than (or in complement to) innovation and renewal. Over the course of her childhood, rather than producing a proper Neo-

[*] First documented by Francis Galton in his 1880 study "Statistics of Mental Imagery," aphantasia refers to the inability to form mental images. See Francis Galton, "Statistics of Mental Imagery." *Mind* 5, no. 19 (1880): 301–18. http://www.jstor.org/stable/2246391.

Victorian subject, the recursive pedagogy of Nell's Primer molds her into a radically empowered insurgent figure, autonomous and defiant in the face of upper-class orthodoxy.

With this, Stephenson offers a speculative allegory for what happens when generative technologies exceed their ideological enclosures. Nell's Primer subverts its intended function because it is *too responsive*: deployed in an uncontrolled environment, it ceases to reinforce legitimacy and instead becomes a generative site of insurgent countervisualization and agency. Later in the novel, this thread reaches its fullest expression in the emergence of the Mouse Army, a vast collective of lower-class girls raised by bootlegged versions of the Primer. Though trained in isolation, over time the girls coalesce into a collective political subject through the shared narratives and affective bonds instilled by their Primers, which, rather than producing fealty to power, have cultivated in them a sense of loyalty to one another.[10] In this sense, the Mouse Army represents an unintended byproduct of generative control: originally intended as a means of naturalizing segregation and elite supremacy, the Primer instead births a powerful insurgency aggregated around its countervisualization.

I contend that the emergence of the fictional Mouse Army gives form to the anxious imaginations surrounding many 'radical' decentralized movements in the real world—what this dissertation has theorized as insurgent aggregation. As such, they point to precisely what is foreclosed by counterinsurgent intensification, representing the potential to visualize subjects, solidarities, and shared political horizons beyond those legible to Visuality 1. Critically, the control dramatized in Stephenson's dystopia is not so much an aberration as a logical extension of the sort embedded in Negroponte's utopian vision of the interface agent, where mediation is made commensurate with the delimitation and modulation of subjects. In *The Diamond Age*, Stephenson offers a critical counterpoint to the disaggregative operations of generativity traced

throughout this dissertation, fictionalizing the very nightmare that 'safe' infrastructures function to avert: a collective, affectively bonded insurgency formed within and against a system designed to disperse it. The challenge, then, is to imagine how generativity might be intentionally redeveloped and redeployed, not to foreclose insurgency in every sense, but to facilitate the visualization of new regimes of legitimacy. Beginning with the recognition of the right to look, this work demands collective appropriations of the technical means through which visualization, narrative, and legitimacy are made operational in networked societies.

Negroponte and Stephenson, though opposing in tone, represent two sides of the same fantasy of generativity as control. But speculation is not only a tool for diagnosing power; it is also indispensable to the work of imagining and realizing its otherwise possibilities. If the speculative mode enables us to trace the arc of power's intensification, it becomes even more vital when turned toward emancipatory ends, though I'd argue that here the bar is considerably higher, insofar as these futures do not arise from presents that readily afford them. On the contrary, present material, political, and discursive conditions either lack—or actively strive to eliminate—the *dispositional*[*] properties from which alternative futures might otherwise emerge. Projects of emancipation must therefore contend not only with the absence of enabling conditions, but with the array of technopolitical forces, tendencies, and inertias inscribing the inevitability of counterinsurgency's intensification. However, despite counterinsurgency's momentum, it is critical to recognize and assert a reality wherein its intensification is not a given—where crisis may yet bring transformation.

---

[*] Futures-studies scholar Wendell Bell uses the term 'dispositional' to describe latent future possibilities. Ending in suffixes like -able, -ible, and -uble, dispositionals mark possible futures as observable characteristics of the present, able to be investigated. In much the same way a vase is always breakable, if never broken, possible futures exist, to varying degrees, as dispositional properties of their present-pasts. See Wendell Bell, *Foundations of Futures Studies: History, Purposes, and Knowledge*, (London: Routledge, 1997): 76.

Indeed, in a reflexive nod to his own work on visuality, Mirzoeff affirms this possibility, noting the naturalization of authority as being undermined by the "simple fact that a counterhistory of Visuality [(such as his)] can be written."[11] That Visuality can be historicized at all marks the crisis of its authority: no longer the unexamined ground, it emerges as an increasingly visible site of contested imagination. Yet, the present conjuncture is unique. Visibility and contestation alone are not enough to unsettle a regime organized around the procedural management of dissent, where the counterinsurgent *requires* resistance to its reality in the form of an "always already potential terrorism"—which it claims the authority to excise—for the perpetual re-legitimation of its power.[12] Here, and with regard to the issue of generative AI, Enzensberger's critique of the bourgeois media system helps clarify the stakes: "It wants the media *as such* and *to no purpose*," he wrote, diagnosing a condition in which the goal of controlling communicative infrastructures persists despite the absence of any narrative or political vision for "making the socially necessary use of them."[13] In the verbiage of this dissertation, counterinsurgency does not so much deploy the media for legitimation, that is, to discursively render its authority "right and hence aesthetic,"[14] but for the 'tactical' operations of classification and separation, now become means and ends. In this sense, the counterinsurgent's aestheticization extends only to the discursive production of its host culture as weak and existentially threatened—requiring permanent intervention and the suspension of politics under the justification of perpetual crisis.

Like the bourgeois media described above, the concept of a "safe" generative media names an infrastructure with potentially unprecedented capacity for visualization, here constrained by technical and political parameters operating under the procedural alibis of safety and epistemic integrity. And so, we might follow Enzensberger (who contrasts his Marxist vision

of an emancipatory media with McLuhan's treatment as autonomous forces shaping perception) in taking seriously Benjamin's call to counter the fascist aestheticization of politics, engaging media through the lens of political struggle, and refashioning generativity as a site of collective countervisualization and political transformation. In claiming generativity as an exercise of the right to look, projects of emancipation must directly engage with the material, social, and technical conditions of how AI models visualize, circumventing the core operations of counterinsurgency to refuse Visuality's separations and invent new forms. This proceeds from the realization—or, rather, reminder—that these technologies are not deterministic, and the tendencies which technical systems come to express as technical *facts* are, in reality, continually oriented and re-oriented through political economic arrangements that are themselves "historical, and perfectly contingent."[15] In *For a New Critique of Political Economy*, Bernard Stiegler emphasizes this contingency, describing technical systems as opening "fields of protentional possibilities"—a largely undifferentiated space from which specific tendencies are selected, stabilized, and later expressed.[16]

This framing is particularly instructive when considering the epistemic politics of generative models, and Stiegler's articulation resonates with Amoore's later description of the models' latent space as a highly generative 'space of possibility' imbued with epistemic and political logics delimiting "what can be known and done in the world."[17] Crucially, Amoore and her co-authors make a point of denying any congruence between generative AI's computational logics and the specific political tendencies expressed by its current manifestations. Instead, they insist that the political emerges in the contingency of the latent space, where positivist claims to the *underlying distribution* are necessarily haunted by the possibility of other, discarded alternatives.[18] This has implications for how we understand, in political terms, the many

'failures' of state-of-the-art generative models. Rather than simply producing 'untrue' outputs, 'hallucinating' models might thus be reimagined as expressing an eruptive surfacing of *statistically plausible* patterns through the model's unsanctioned amplification of weak signals—fragments of alternative distributions heretofore excluded, suppressed, or rendered unintelligible.

To engage this critically is not so much to mistake hallucination for visualization, but to recognize in such misfires the inchoate outlines of otherwise possibility—alternatives that persist in the latent space as an affective excess, irreducible to pattern, but nonetheless present as a "potentially transformative charge."[19] The creative potentiality of latency is made explicit in the work of sound artists like Holly Herndon and Mat Dryhurst who, by directing generative models to create tracks that are intentionally 'noisy' or divergent from the statistically-derived codes of the music they were trained on, in effect 'mine' the models' space in the hopes of unearthing new aesthetic forms in the form of what would otherwise be considered failure.[20] Recalling the literature review, where—likening it to the "blank spaces of the map" evoked in Conrad's *Heart of Darkness*—I presented a framing of the model's interstitial vector space as the 'blind spot' of visuality, the latent space now emerges as a site of wonder, exploration, and world-making. In political terms, the task of countervisualization thus starts with cultivating the unrealized, suppressed, or overlooked affordances of generativity and shifting, as Stiegler proposes, from the resigned inevitability of *there is no alternative* to a politics of *there are lots of alternatives*—leveraging generativity not to foreclose futures, but to reopen them.

This requires a rethinking of the epistemic and political affordances of generative technologies. Writing on the effects of synthetic media on testimonial practice, social philosopher Megan Hyska identifies a key paradox: as much as generative AI erode the epistemic value of recordings and thus decrease our ability to convey information that a receiver

understands to be independent of persuasive intent—what she calls acts of *showing*—they simultaneously expand our capacity to communicate *with* informative or persuasive intent by enabling new acts of multimodal *telling*.[21] Contrary to popular understandings, telling in this context needn't equate with deception, and rather than misrepresent the past or present, these technologies may afford means to tell one another, in vivid terms, about our visions for the future.[22] In this way, generative AI partially redistribute the communicative power of media from the domain of *past verification* to that of *future imagination*.[23] While Hyska is ultimately skeptical about this shift, claiming that it unsettles the "relational equality" between communicating parties (which she sees as critical to the sustainability of collective action), such a concern needn't foreclose the utility of generative media to projects of countervisualization, particularly where the goal is not so much consensus, as taking a first step in unsettling social and political orthodoxy through the visualization of alternatives.

This orientation toward the plural and the possible similarly animates the work of Dutch artist Jonas Staal, whose concept and practice of a 'propaganda arts' rejects deception and top-down media control in favour of an emancipatory model grounded in the creation of "collaborative propaganda interfaces," a term he uses to describe tools through which the masses can actively participate in the co-construction of their realities through art.[24] Set in opposition to what Staal calls *elite* propaganda, these *popular* propagandas perform an inversion of visuality, liberating us from what we "think the world is"—that is to say, the world as narrated by power—through the re-appropriation of visualization as a means of imagining "what we want it to become."[25] It follows, then, that generative media may offer new affordances for the collective performance of narrative power through multimodal acts of *telling*, here refigured as acts of "visualizing, staging, and performing the new realities that propaganda brings about."[26]

By providing new tools and infrastructures for partaking in the processes of world-making, popular propagandas effectively transform receivers into producers, enlisting individuals as the "micro-performative vessels" of the ideological motives that sustain the propaganda effort.[27] In this way, Staal echoes Enzensberger, whose vision of an emancipatory media politics stemmed from a belief that the contradiction between the role of producers and consumers is not inherent to the media, but externally enforced,[*] and that the horizontal tendencies of what he called "the electronic media" could only be made "fully productive" in a truly free and egalitarian society.[28] Rejecting the fantasy of an unmanipulated or neutral media, Enzensberger insisted that an emancipatory strategy, rather than striving to eliminate manipulation, must "make *everyone* a manipulator"[29]—an imperative that finds contemporary expression in the collective visualization and performance of alternatives evoked by Staal. What Enzensberger imagined as a decentralized, self-organizing media ecology reappears here in the collaborative interfaces of popular propaganda, where media manipulation is no longer a mark of domination, but of *contribution*,[†] extending the invitation to co-author the visual and discursive conventions that shape social and political life. It is becoming increasingly common to lament the proliferation of AI-generated 'slop,' the term assigned to the deluge of low-quality, unnecessary, or incoherent generative outputs said to clutter content feeds and dilute public discourse. While I do not deny the existence or inconvenience of genuine slop, I want to suggest

[*] According to Enzensberger, the distinction between receivers and transmitters, consumers and producers, ultimately mirrors the "basic contradiction between the ruling class and the ruled class." See Enzensberger, "Constituents of a Theory of the Media,": 15.

[†] Bernard Stiegler contrasts the "economy of contribution" with what he calls the "economy of carelessness," the latter defined by passive consumption, algorithmic individuation, and the erosion of collective bonds. An economy of contribution, on the other hand, seeks to re-engage individuals as active participants in the production of shared meaning and symbolic life—restoring agency through practices of co-creation and epistemic investment. See Bernard Stiegler, *For a New Critique of Political Economy* (Polity, 2010), 129.

that the sort of noise it represents is not without political and epistemic significance insofar as unbounded generation, in its sloppy excess, might instead be a sort of *primordial goo*—a chaotic substrate from which new forms of life and living emerge.

In this light, generative media could offer real utility to projects of emancipatory countervisualization. While this dissertation has charted their development as agents of counterinsurgent intensification through the technical enforcement of separation, the very capacities that make them effective instruments of control may also render them uniquely suited to projects of collective narration and sensemaking. Beyond rendering visualization as a programmable (and therefore mutable) operation, generative media unsettle the presumed neutrality of representation in another sense. By democratizing access to the epistemic credibility traditionally reserved for the visible and the verified, generative models allow users to produce media with the aesthetic and rhetorical weight afforded to realism, even when visualizing in the absence of a corresponding *present* reality. Contra the framing of this capability as simulation or deceit, emancipatory propagandas might thus deploy the model as a powerful speculative engine: capable of vividly illustrating the collective imagination of worlds and relations that don't yet exist, but should. To manipulate, then, is to stake a claim to the right to look—an exercise of autonomy in arranging 'the relations of the visible and the sayable' through intervention in the protocols of a visuality that has long denied it.

As evidenced by growing interest in repurposing generative models for creative ends, this vision is not strictly theoretical. For example, the work of Pasquier et al. (affiliated with the Metacreation Lab for Creative AI at SFU) focuses on developing tools and frameworks enabling artists to work with large (and small) generative models, which they categorize into three modes of intervention: *navigating* existing generative space; *adapting* models through fine-tuning; and

crafting new and existing model architectures to produce generative spaces that are "qualitatively different" from existing options.[30] Crucially, this project aims to lower technical barriers so that artists and other non-expert users can become what they call "model crafters"—emancipated manipulators newly empowered with the agency to 'seize' the means of visualization. These practices are not just acts of artistic experimentation, they are direct and deliberate interventions in the procedural logics of generativity to foster what is called *active divergence*: directing models away from target distributions to visualize novel or otherwise 'out-of-distribution' patterns. In the present context, projects like this open important pathways for thinking the adaptation and crafting of generative models as *propagandas* in their own right: concerted efforts to rearrange or erode the fixed features of models' epistemic landscape so that we might visualize realities that "overflow and exceed their narrowly prescribed boundaries of who or what can count."[31]

According to Staal's formulation, the invigoration of a popular propaganda arts is central to the practical realization of Judith Butler's theory of *performative assembly*, understood as the simultaneous enactment of a collective subjectivity in otherwise contested space and formalized in the practice of "assemblism," which is itself a propaganda.[32] It follows, then, that the emancipatory potential of media lay not just in the ability to create art, but to enable the disaggregated masses' *self-visualization as a people*. Stephenson's narrative of the Mouse Army offers a speculative image of the role that generativity might play in this process, presenting the insurgency's emergence as an effect of the models' fostering of affective solidarity around countervisualizations produced through divergence, both active and accidental. This sort of assembly is precisely the object of an intensified counterinsurgent visuality. Without pretense to

discipline, reform, or placation of the disaffected, it seeks to destroy the connections that might otherwise bind them into politically legible collectives.

Strategies of disaggregation do not preclude insurgency, but atomize it—indeed, the counterinsurgent thrives on and maintains difference as the incoherent 'noise' against which its legitimacy is defined in crisis. This development runs alongside two critical inversions: both in how mainstream or fringe positions are defined, and in how visuality imagines the temporality of its counters. To the first point, the ability to assess sentiment and general adherence in the population has lead to the realization that ideological and cultural positions once visualized as statistical outliers are often quantitatively more popular than what was, and still is, imagined as the 'mainstream.' No longer able to deny it, authority now internalizes this 'new' reality, fortifying its border, the last bastion of legitimacy against the insurgents at the gate. Whereas the imperial visualizer imagined its other as historically anterior, a backward subject to be either excluded or made modern, authority now visualizes itself as holding back an array of imagined futures where the insurgent has won out. So intent on stabilizing a present that escapes its control, the counterinsurgent meets the future with a nostalgic desire for its own imagined past.

In the ongoing struggle, those interested in emancipation must not only ask who authors reality in our name but also strive to control the means by which that reality is visualized and governed.[33] Generative models, as this dissertation has shown, are rapidly becoming central to that visualization. As much as their plasticity makes them dangerous, it is also potentially generative insofar as they expand the affordances through which we might challenge the "inertial, undead defaults" of Visuality 1. The challenge ahead is to ensure that these expanded capacities are not monopolized by elite interests or bent toward counterinsurgent control, but remain available to the right to look. The widespread discontents which animate the ongoing

crisis of establishment legitimacy require more places to go than underground or further into the arms of right-wing populists. They demand new and intensified means of visualizing alternatives, so that we may assemble new realities worthy of inheriting authority from the dead and dying remains of the present. This is the emancipatory wager of countervisualization: not only that generative models can help us visualize differently, but that through them, we might begin to visualize *together*.

With this dissertation, I have claimed that the 'safe' generative model both produces—and is the product of—a reactionary intensification of counterinsurgent visuality by a neoliberal establishment in crisis. To diagnose this intensification across historical, political, and technical registers, this work proceeded in three parts: a genealogical account of media, visuality, and power; an analysis of crisis, disinformation discourse, and aggregated insurgency as drivers of intensification; and a study of the regulatory, technical, and epistemic closures producing the generative model as an agent of intensified counterinsurgent control. I have argued that so-called 'safe' models threaten to ensconce disaggregation in the technical infrastructures of knowledge and communication, effectively rendering counterinsurgency "a regular function, coextensive with society,"[34] and policing the increasingly contested legitimacy of political and epistemic authority.

While this dissertation offers a critical account of establishment institutions—in particular their securitization of generative AI via discourses of disinformation, safety, and existential risk—it should not be mistaken as an argument broadly in favour of deregulation. Even in the speculative turn toward emancipated visualization, this project ultimately calls not for the absence of governance, but for its relegitimation through reorientation. Indeed, my critique is directed at the broader neoliberal condition wherein the deregulation of capital runs

parallel to the intensified regulation of increasingly disaggregated and dispossessed populations. The development and deployment of 'safe' generativity as a balm for the risks of rapid technological and economic growth reflects this dynamic: as data extraction and market concentration proceed largely unimpeded, the management of online speech, assembly, and narrative escalates under new rhetorical and technical guises. As such, this work does not reject the regulatory principle but critiques its implementation as a mechanism of counterinsurgent delegitimation and control, following others in calling for the reclamation and strengthening of media and communications as critical sites of democratic deliberation and political reimagination. Moving in this direction would require that establishment institutions reckon with their role in producing the present crisis and undertake the long and difficult task of building new political economies organized around responsiveness rather than control. What is called for is not simply more or less regulation, but new forms of governance that are substantively democratic, materially accountable, and capable of supporting the imaginative labour required to visualize and assemble otherwise futures.

By effectively diagnosing the opposite, it is my hope that this dissertation will be of use to future work concerned with reorienting generative AI toward emancipatory ends. If dominant institutions have been successful in securitizing the media as a threat to stability, I hope this project may be understood as an act of *counter*-securitization, framing the intensification of counterinsurgent visuality through 'safe' generativity as a greater and more fundamental threat to core democratic values than those presented by disinformation and epistemic chaos—both real and imagined. As I look ahead to postdoctoral work, I plan to extend this research by collaborating with artists, activists, and critical technologists to explore methods for circumventing the stultifying epistemic and political effects of counterinsurgent visualization

through interventions in the structure, operation, and deployment of generativity as an exercise in worldmaking.

We are at a significant political and technological juncture. If the hegemony of the counterinsurgent mode of governance was consolidated with the "end of history"—so marked by the presumed finality of liberal legitimacy—then its intensification marks history's implacable return. The moral panics around disinformation and AI, though reactionary in most respects, have been productive insofar as they've extended politics into infrastructures that have, to this point, largely evaded serious public scrutiny.[35] As Langdon Winner reminds us, it is in the earliest stages of a technology's development, before its structures and conventions congeal, that the greatest latitude of choice exists. Rather than dismiss the present as a moment of democratic unraveling, we might choose to see it as one of great possibility: a rupture from which new forms might yet emerge. If counterinsurgency names a condition where 'nothing ever happens,' every deviation already foreclosed, then our task is to pry this fissure wide open, until "from a situation in which nothing can happen, suddenly anything is possible again."[36]

ENDNOTES

INTRODUCTION

[1] Jessica Riskin, "The Defecating Duck, or, the Ambiguous Origins of Artificial Life," *Critical Inquiry* 29, no. 4 (June 2003): 599.

[2] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation* 18, no. 7 (July 2006): 1527–54.

[3] R. Salakhutdinov and Geoffrey E. Hinton, "Deep Boltzmann Machines," in *International Conference on Artificial Intelligence and Statistics*, 2009.

[4] Yoshua Bengio et al., "Deep Generative Stochastic Networks Trainable by Backprop," in *Proceedings of the 31st International Conference on Machine Learning* (International Conference on Machine Learning, PMLR, 2014), 226–34.

[5] Ian J. Goodfellow et al., "Generative Adversarial Networks" (arXiv, June 10, 2014): 2.

[6] Ibid.

[7] "GPT-4o System Card," OpenAI, accessed November 28, 2024, https://openai.com/index/gpt-4o-system-card/.

[8] Scott Mulligan, "AI Trained on AI Garbage Spits out AI Garbage," MIT Technology Review, July 24, 2024, https://www.technologyreview.com/2024/07/24/1095263/ai-that-feeds-on-a-diet-of-ai-garbage-ends-up-spitting-out-nonsense/.

[9] Jocelyn Mintz, "'Garbage in, Garbage out': AI Fails to Debunk Disinformation, Study Finds," Voice of America, October 21, 2024.

[10] "A 'Super Year' for Elections," United Nations Development Programme, accessed November 26, 2024, https://www.undp.org/super-year-elections.

[11] Barbara A. Trish, "4 Ways AI Can Be Used and Abused in the 2024 Election, from Deepfakes to Foreign Interference," The Conversation, October 16, 2024.; Pooja Chhabria, "The Big Election Year: How to Protect Democracy in the Era of AI," World Economic Forum, January 29, 2024,; Vittoria Elliot, "2024 Is the Year of the Generative AI Election," WIRED, May 30, 2024.

[12] Di Cooke, "Synthetic Media and Election Integrity: Defending Our Democracies" (Centre for Emerging Technology and Security, August 2023).

[13] "AI's Impact on 2024 Global Elections," Aspen Digital, accessed November 26, 2024, https://www.aspendigital.org/event/ai-2024-global-elections/.

[14] Shanay Gracia, "Americans in Both Parties Are Concerned over the Impact of AI on the 2024 Presidential Campaign," *Pew Research Center* (blog), September 19, 2024, https://www.pewresearch.org/short-reads/2024/09/19/concern-over-the-impact-of-ai-on-2024-presidential-campaign/.

[15] Ethan Mesquita et al., "Preparing for Generative AI in the 2024 Election: Recommendations and Best Practices Based on Academic Research," White paper (Chicago: University of Chicago Harris School of Public Policy, August 2023): 5.

[16] Alexander Puutio and David A. Timis, "Deepfake Democracy: Here's How Modern Elections Could Be Decided by Fake News," World Economic Forum, October 5, 2020, https://www.weforum.org/stories/2020/10/deepfake-democracy-could-modern-elections-fall-prey-to-fiction/.

[17] Mark Labbe, "The Deepfake 2020 Election Threat Is Real, but Containable," TechTarget, September 25, 2020, https://www.techtarget.com/searchenterpriseai/feature/The-deepfake-2020-election-threat-is-real-but-containable.

[18] OpenAI, "How OpenAI Is Approaching 2024 Worldwide Elections," accessed November 26, 2024, https://openai.com/index/how-openai-is-approaching-2024-worldwide-elections/.

[19] Andrew R. Chow, "AI's Underwhelming Impact On the 2024 Elections," TIME, October 30, 2024, https://time.com/7131271/ai-2024-elections/; Pranshu Verma, Will Oremus, and Cat Zakrzewski, "AI Didn't Sway the Election, but It's Eroding Voters' Grip on Reality," The Washington Post, November 9, 2024, https://www.washingtonpost.com/technology/2024/11/09/ai-deepfakes-us-election/.

[20] Douglas Porpora and Seif Sekalala, "Truth, Communication, and Democracy," *International Journal of Communication* 13, no. 0 (February 26, 2019).

[21] Di Cooke et al., "Crossing the Deepfake Rubicon: The Maturing Synthetic Media Threat Landscape" (Center for Strategic and International Studies (CSIS), 2024): 2.

[22] Ibid., 3.

[23] Jared Cohen and George Lee, "The Generative World Order: AI, Geopolitics, and Power," Goldman Sachs Insights, December 14, 2023, https://www.goldmansachs.com/insights/articles/the-generative-world-order-ai-geopolitics-and-power.

[24] Nate Soares and Benja Fallenstein, "Aligning Superintelligence with Human Interests: A Technical Research Agenda" (Machine Intelligence Research Institute, 2014).; Jan Leike et al., "Scalable Agent Alignment via Reward Modeling: A Research Direction" (arXiv, November 19, 2018)

[25] Nathalie A. Smuha, "Beyond the Individual: Governing AI's Societal Harm," *Internet Policy Review* 10, no. 3 (September 30, 2021), https://policyreview.info/articles/analysis/beyond-individual-governing-ais-societal-harm.

[26] Tao Feng et al., "How Far Are We From AGI: Are LLMs All We Need?" (arXiv, November 24, 2024), https://doi.org/10.48550/arXiv.2405.10313.

[27] Irene Solaiman et al., "Evaluating the Social Impact of Generative AI Systems in Systems and Society" (arXiv, June 28, 2024), https://doi.org/10.48550/arXiv.2306.05949.

[28] Adrienne Thompson, "The Existential Threat of AI-Enhanced Disinformation Operations," C*enter for Security and Emerging Technology* (blog), July 11, 2022, https://cset.georgetown.edu/article/the-existential-threat-of-ai-enhanced-disinformation-operations/

[29] Nina Schick, "Fucked-up dystopia," in *Deepfakes: The Coming Infocalypse*, Ebook: para. 8-11.

[30] Elizabeth Seger, "Should Epistemic Security Be a Priority GCR Cause Area," *Intersections, Reinforcements, Cascades*, 2023: 19.

[31] Mark Fisher, *Capitalist Realism: Is There No Alternative?* (Zero Books, 2009): 78.

[32] Ganesh Sitaraman, "The Collapse of Neoliberalism," *The New Republic*, December 23, 2019, https://newrepublic.com/article/155970/collapse-neoliberalism.

[33] NPR Staff, "Occupy Wall Street Inspires Worldwide Protests," *NPR*, October 15, 2011, sec. World, https://www.npr.org/2011/10/15/141382468/occupy-wall-street-inspires-worldwide-protests.

[34] Chenda Ngak, "Occupy Wall Street Uses Social Media to Spread Nationwide - CBS News," October 13, 2011, https://www.cbsnews.com/news/occupy-wall-street-uses-social-media-to-spread-nationwide/.

[35] Anastasia Kavada, "Creating the Collective: Social Media, the Occupy Movement and Its Constitution as a Collective Actor," *Information, Communication & Society* 18, no. 8 (August 3, 2015): 872–86, https://doi.org/10.1080/1369118X.2015.1043318.

[36] Lizzy Davies, "Occupy Movement: City-by-City Police Crackdowns so Far," *The Guardian*, November 15, 2011, sec. US news, https://www.theguardian.com/world/blog/2011/nov/15/occupy-movement-police-crackdowns.

[37] Edward S. Herman and Noam Chomsky, *Manufacturing Consent: The Political Economy of the Mass Media.*, (Pantheon Books: 1989).

[38] Wendy Hui Kyong Chun and Alex H. Barnett, *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition* (Cambridge, Massachusetts London, England: The MIT Press, 2021), 82.

[39] Britt Paris, "Seeing Through the Fog of War: Assessing Epistemic Burden Around Cheapfakes and Deepfakes of Geopolitical Crisis," in *Re-Thinking Mediations of Post-Truth Politics and Trust* (Routledge, 2023), 154.

[40] "Oxford Word of the Year 2016," Oxford Languages, accessed December 6, 2024, https://languages.oup.com/word-of-the-year/2016/.

[41] Andreas Jungherr and Ralph Schroeder, "Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy," *Social Media + Society* 7, no. 1 (January 1, 2021): 2.

[42] Farzad Shahinfard, "The Battle against AI-Driven Disinformation," Policy Options, September 30, 2024, https://policyoptions.irpp.org/magazines/september-2024/defending-canada-the-battle-against-ai-driven-disinformation/.

[43] Sascha Brodsky, "Tech Industry Ramps up Efforts to Combat Rising Deepfake Threats," *IBM Blog* (blog), September 10, 2024, https://www.ibm.com/blog/deepfake-detection/.

[44] Jack Aldane, "Tech Firms Sign Pact to Counter AI Deepfakes as Elections near - Global Government Forum," Global Government Forum, February 19, 2024, https://www.globalgovernmentforum.com/tech-firms-sign-pact-to-counter-ai-deepfakes-as-elections-near/.

[45] Théophile Lenoir and Chris Anderson, "Introduction Essay: What Comes After Disinformation Studies," *Center for Information, Technology, & Public Life (CITAP)*, University of North Carolina at Chapel Hill, January 23, 2023.

[46] Cathy Li and Agustina Callegari, "How AI Can Also Be Used to Combat Online Disinformation," World Economic Forum, June 14, 2024, https://www.weforum.org/stories/2024/06/ai-combat-online-misinformation-disinformation/.

[47] Alexei Abrahams and Gabrielle Lim, "Repress/Redress: What the 'War on Terror' Can Teach Us about Fighting Misinformation," *Harvard Kennedy School Misinformation Review*, July 22, 2020: 1.

[48] Alexei Abrahams and Gabrielle Lim, "Repress/Redress […]": ibid.

[49] David Kilcullen, *Counterinsurgency* (Oxford University Press, USA, 2010): 1.

[50] Peter Sterling, "Allostasis: A Model of Predictive Regulation," Physiology & Behavior 106, no. 1 (April 12, 2012): 5–15.

[51] Andrea Bardin and Marco Ferrari, "Governing Progress: From Cybernetic Homeostasis to Simondon's Politics of Metastability," *The Sociological Review* 70, no. 2 (March 2022): 255.

[52] Gilbert Simondon, *On the Mode of Existence of Technical Objects* (Univocal Publishing, 2017): 162.

[53] Koyré qtd. In Bardin and Ferrari, "Governing Progress […]," 249.

[54] Jack Bratich, "Civil Society Must Be Defended: Misinformation, Moral Panics, and Wars of Restoration," *Communication, Culture and Critique* 13, no. 3 (September 1, 2020): 314.

[55] Alex Hern, "Facebook: We Were Too Slow to Recognise Our 'corrosive' Effect on Democracy," *The Guardian*, January 22, 2018, sec. Technology, https://www.theguardian.com/technology/2018/jan/22/facebook-too-slow-social-media-fake-news-hiring.

[56] Olivia Solon, "Facebook's Fake News: Mark Zuckerberg Rejects 'crazy Idea' That It Swayed Voters," *The Guardian*, November 11, 2016, sec. Technology, https://www.theguardian.com/technology/2016/nov/10/facebook-fake-news-us-election-mark-zuckerberg-donald-trump.

[57] "Hard Questions: What's Facebook's Strategy for Stopping False News?," *Meta* (blog), May 23, 2018, https://about.fb.com/news/2018/05/hard-questions-false-news/.

[58] X Safety, "Freedom of Speech, Not Reach: An Update on Our Enforcement Philosophy," Blog, *X Safety* (blog), April 17, 2023, https://blog.x.com/en_us/topics/product/2023/freedom-of-speech-not-reach-an-update-on-our-enforcement-philosophy.

[59] Monika Bickert, "Our Approach to Labeling AI-Generated Content and Manipulated Media," *Meta* (blog), April 5, 2024, https://about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/.

[60] Bratich, "Civil Society Must Be Defended […]": 319. [emphasis added]

[61] "Moderations" OpenAI Platform, accessed April 8, 2024, https://platform.openai.com./docs/api-reference/moderation

[62] "How OpenAI Is Approaching 2024 Worldwide Elections," OpenAI, January 15, 2024, https://openai.com/index/how-openai-is-approaching-2024-worldwide-elections/.

[63] "Preparing for Global Elections in 2024," Anthropic, February 16, 2024, https://www.anthropic.com/news/preparing-for-global-elections-in-2024.

[64] Susan Jasper, "How We're Approaching the 2024 U.S. Elections," *The Keyword* (blog), December 19, 2023, https://blog.google/outreach-initiatives/civics/how-were-approaching-the-2024-us-elections/.

[65] "Safeguards for Elections," Meta, https://about.meta.com/actions/preparing-for-elections-with-meta

[66] Anh Bui et al., "Erasing Undesirable Concepts in Diffusion Models with Adversarial Preservation" (arXiv, October 29, 2024), https://doi.org/10.48550/arXiv.2410.15618.

[67] Michael Chui et al., "NOTES FROM THE AI FRONTIER: APPLYING AI FOR SOCIAL GOOD," Discussion Paper (McKinsey Global Institute, December 2018), https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.pdf.

[68] Garance Burke, "Researchers Say an AI-Powered Transcription Tool Used in Hospitals Invents Things No One Ever Said," AP News, October 26, 2024, https://apnews.com/article/ai-artificial-intelligence-health-business-90020cdf5fa16c79ca2e5b6c4c9bbb14.

[69] Barry Buzan, Ole Wæver, and Jaap de Wilde, *Security: A New Framework for Analysis* (Lynne Rienner Publishers, 1998): 23.

[70] Mathew Ingram, "The Supreme Court Hears Arguments about Government 'Jawboning,'" Columbia Journalism Review, accessed December 14, 2024, https://www.cjr.org/the_media_today/supreme_court_louisiana_missouri_social_media.php.

[71] W. Lance Bennett and Steven Livingston, "A Brief History of the Disinformation Age: Information Wars and the Decline of Institutional Authority," in *The Disinformation Age*, ed. Steven Livingston and W. Lance Bennett, SSRC Anxieties of Democracy (Cambridge: Cambridge University Press, 2020), 3–40.

[72] Nicholas Mirzoeff, *The Right to Look: A Counterhistory of Visuality* (Durham: Duke University Press, 2011).

[73] Ibid., 3.

[74] Ibid., 5.

[75] Ibid., 24.

[76] Alexander R. Galloway, *Protocol: How Control Exists after Decentralization*, Leonardo (Cambridge, Mass: MIT Press, 2004): 75. [emphasis original]

[77] Ibid., 52.

[78] Michel Foucault, *Society Must Be Defended: Lectures at the College de France, 1975-76*, 1st ed. (Picador, 2003): 16.

[79] Wendy Chun qtd. in Sonja Solomun, "Polarization as the Technological Goal – Not the Error," Centre for Media, Technology and Democracy, accessed August 6, 2024, https://www.mediatechdemocracy.com/all-work/polarization-as-the-technological-goal-not-the-error

[80] Barry Buzan, Ole Wæver, and Jaap de Wilde, *Security: A New Framework for Analysis* (Lynne Rienner Publishers, 1998): 24.

[81] Christopher St Aubin and Jacob Liedke, "-Most Americans Favor Restrictions on False Information, Violent Content Online," Pew Research Center (blog), July 20, 2023, https://www.pewresearch.org/short-reads/2023/07/20/most-americans-favor-restrictions-on-false-information-violent-content-online/

[82] Felix M. Simon, Sacha Altay, and Hugo Mercier, "Misinformation Reloaded? Fears about the Impact of Generative AI on Misinformation Are Overblown," *Harvard Kennedy School Misinformation Review*, October 18, 2023: 4. https://doi.org/10.37016/mr-2020-127.

[83] Alberto Acerbi, Sacha Altay, and Hugo Mercier, "Research Note: Fighting Misinformation or Fighting for Information?," *Harvard Kennedy School Misinformation Review*, January 12, 2022, https://doi.org/10.37016/mr-2020-87.

[84] Michel Foucault, *Discipline and Punish: The Birth of the Prison*. Trans. Alan Sheridan (New York: Vintage, 1979): 81.

[85] Kilcullen, *Counterinsurgency*: 194.

[86] Ibid., 42.

[87] "List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words," GitHub, accessed April 2, 2024, https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words/blob/master/en.

[88] Chinese National Information Security Standardization Technical Committee, Ben Murphy trans. *Basic Safety Requirements for Generative Artificial Intelligence Services (Draft for Feedback)*, Oct. 2023, https://cset.georgetown.edu/publication/china-safety-requirements-for-generative-ai/

[89] "Membership - C2PA," Coalition for Content Provenance and Authenticity, accessed January 15, 2025, https://c2pa.org/membership/.

[90] "Christchurch Call" Christchurch Call, accessed May 28, 2024, https://www.christchurchcall.com

[91] "Terrorist Content Analytics Platform," Terrorist Content Analytics Platform, accessed January 16, 2025, https://www.terrorismanalytics.org/.

[92] Ronald Watkins, "USSOCOM to Use AI to Detect Disinformation Threats on Social Media," The Defense Post, August 31, 2023, https://www.thedefensepost.com/2023/08/31/ussocom-ai-disinformation-social-media/.

[93] Alexander R. Galloway, *Gaming: Essays on Algorithmic Culture*, Electronic Mediations ; v. 18 (Minneapolis: University of Minnesota Press, 2006), 63.

[94] Mirzoeff, *The Right to Look*, 1.

[95] Foucault, *Discipline and Punis*h, 82.

[96] Mirzoeff, *The Right to Look* […]: 26.

LITERATURE REVIEW

[1] Martin Staniland, *What Is Political Economy?: A Study of Social Theory and Underdevelopment* (New Haven: Yale University Press, 1985): 10.

[2] Several accounts mark the emergence and development of a political economy of media. For example, Edward L. Bernays, *Propaganda: The Public Mind in the Making*. Liveright, 1936; Harold D. Lasswell, "The Theory of Political Propaganda," *The American Political Science Review* 21, no. 3 (1927): 627–31; Harold Adams Innis, *Empire and Communications* (Clarendon Press, 1950); Theodor W. Adorno and Max Horkheimer. "The Culture Industry: Enlightenment as Mass Deception." In *Dialectic of Enlightenment: Philosophical Fragments*. Edited by Gunzelin Schmid Noerr and translated by Edmund Jephcott. Stanford, CA: University of Stanford Press, 2002: 94–136; Graham Murdock and Peter Golding, "For a Political Economy of Mass Communications," *Socialist Register* 10 (March 18, 1973); Vincent Mosco. *The Political Economy of Communication: Rethinking and Renewal*. SAGE Publications, 1996.

[3] Hannah Barker, *Newspapers, Politics and English Society, 1695-1855* (Longman, 2000): 225.

[4] James Curran, foreword to *Critical Political Economy of the Media: An Introduction*, by Jonathan Hardy, 1st ed., Communication and Society (Routledge, 2014): xiv.

[5] Edward S. Herman and Noam Chomsky, *Manufacturing Consent: The Political Economy of the Mass Media* (Pantheon Books, 1988): 2.

[6] David Karpf, *The MoveOn Effect: The Unexpected Transformation of American Political Advocacy* (Oxford University Press, 2012): 10.

[7] Tarleton Gillespie, "The Relevance of Algorithms," in *Media Technologies: Essays on Communication, Materiality, and Society*, 1st ed., Inside Technology (The MIT Press, 2014): 169. [emphasis added]

[8] Ibid.

[9] Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven, CT: Yale University Press, 2008): 60.

[10] David M. Berry, "The Poverty of Networks," *Theory, Culture & Society* 25, no. 7–8 (December 1, 2008): 369.

[11] Christian Fuchs, *Social Media: A Critical Introduction*, 1st ed. (SAGE Publications Ltd, 2013): 102.

[12] Ibid., 105.

[13] Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (PublicAffairs, 2019): 324.

[14] Ibid., 353.

[15] Tarleton Gillespie, "The Relevance of Algorithms,": 168.

[16] David M Berry, "The Relevance of Understanding Code to International Political Economy," *International Politics* 49, no. 2 (March 1, 2012): 278.

[17] Langdon Winner, "Do Artifacts Have Politics?," *Daedalus* 109, no. 1 (1980): 127.

[18] Ibid., 124.

[19] Ibid.

[20] Ibid., 128.

[21] Lawrence Lessig, *Code: And Other Laws of Cyberspace, Version 2.0* (Basic Books, 2006): 5.

[22] Wendy Hui-Kyong Chun, *Control and Freedom: Power and Paranoia in the Age of Fiber Optics* (Cambridge, Mass: MIT Press, 2006): 66-67.

[23] Alexander R. Galloway, *Protocol: How Control Exists after Decentralization* (Cambridge, Mass: MIT Press, 2004): 10.

[24] Ibid., 8.

[25] Winner, "Do Artifacts Have Politics?,": 130.

[26] Ibid., 128.

[27] Adrian Mackenzie, *Machine Learners: Archaeology of a Data Practice*, MIT Press (The MIT Press, 2017).

[28] Pei Wang, "On Defining Artificial Intelligence," *Journal of Artificial General Intelligence* 10, no. 2 (January 1, 2019): 1–37.

[29] Dagmar Monett and Colin W. P. Lewis, "Getting Clarity by Defining Artificial Intelligence—A Survey," ed. Vincent C. Müller, vol. 44, *Studies in Applied Philosophy, Epistemology and Rational Ethics* (Philosophy and Theory of Artificial Intelligence 2017, Berlin: Springer International Publishing, 2018), 212–14.

[30] José Hernández-Orallo, *The Measure of All Minds: Evaluating Natural and Artificial Intelligence* (Cambridge: Cambridge University Press, 2017): 152.

[31] Lucy Suchman, "Six Unexamined Premises Regarding Artificial Intelligence and National Security," *AI Now Institute* (blog), March 31, 2021, https://ainowinstitute.org/publication/six-unexamined-premises-regarding-artificial-intelligence-and-national-security.

[32] Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Third edition, Global edition, Prentice Hall Series in Artificial Intelligence (Boston Columbus Indianapolis: Pearson, 2016): 34.

[33] Ibid., 4. [emphasis added]

[34] Ibid., 36-37.

[35] John Haugeland, *Artificial Intelligence: The Very Idea*, 1st ed. (MIT Press, 1985): 113.

[36] José Mira, "Symbols versus Connections: 50 Years of Artificial Intelligence," *Neurocomputing*, Neural Networks: Algorithms and Applications, 71, no. 4 (January 1, 2008): 675.

[37] Geoffrey E. Hinton, "Connectionist Learning Procedures," *Artificial Intelligence* 40, no. 1 (September 1, 1989): 186.

[38] Daniel Crevier, *AI: The Tumultuous History Of The Search For Artificial Intelligence* (Basic Books, 1993); Nils J. Nilsson, *The Quest for Artificial Intelligence* (Cambridge University Press, 2009); Michael Wooldridge, *A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going* (Flatiron Books, 2021).

[39] Maximilian Kasy, "The Political Economy of AI: Towards Democratic Control of the Means of Prediction," Discussion Paper Series (IZA Institute of Labour Economics, April 2024): 3.

[40] Taina Bucher, *If...Then: Algorithmic Power and Politics*, Oxford Studies in Digital Politics (Oxford University Press, USA, 2018): 41.

[41] Bruno Latour, *Pandora's Hope: Essays on the Reality of Science Studies*, 2. print (Cambridge, Mass.: Harvard University Press, 2000): 304.

[42] Latour, *Pandora's Hope* […]: 183.

[43] Frank Pasquale, "Introduction: The Need to Know," in *The Black Box Society*, The Secret Algorithms That Control Money and Information (Harvard University Press, 2015), 2.

[44] Ibid., 3 & 8.

[45] Ibid., 9.

[46] John D. Kelleher, *Deep Learning*, The MIT Press Essential Knowledge Series (Mit Press, 2019): 245.

[47] Pierre Azoulay, Joshua L. Krieger, and Abhishek Nagaraj, "Old Moats for New Models: Openness, Control, and Competition in Generative AI," Working Paper, Working Paper Series (National Bureau of Economic Research, May 2024);

[48] M. Beatrice Fazi, "Beyond Human: Deep Learning, Explainability and Representation," *Theory, Culture & Society* 38, no. 7–8 (December 2021): 63.

[49] Ibid., 58 & 62.

[50] Will Heaven, "Large Language Models Can Do Jaw-Dropping Things. But Nobody Knows Exactly Why.," MIT Technology Review, March 4, 2024.

[51] Jason Wei et al., "Emergent Abilities of Large Language Models" (arXiv, October 26, 2022)

[52] See also: Shane Denson, "Discorrelation and Post-Cinema," in *Discorrelated Images* (Durham, NC: Duke University Press, 2020); Adrian MacKenzie and Anna Munster, "Platform Seeing: Image Ensembles and Their Invisualities," *Theory, Culture & Society* 36, no. 5 (September 1, 2019): 3–22, among others.

[53] Arun Rai, "Explainable AI: From Black Box to Glass Box," *Journal of the Academy of Marketing Science* 48, no. 1 (January 1, 2020): 137–41.

[54] Taina Bucher, *If...Then* […]: 47.

[55] Bruno Latour, *Reassembling the Social. An Introduction to Actor-Network-Theory* (Oxford University Press, 2005).

[56] Taina Bucher, *If...Then* […]: 49.

[57] Ibid., 55.

[58] Jonathan Roberge and Michael Castelle, eds., *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies* (Cham: Springer International Publishing, 2021): 5.

[59] Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Third edition, Global edition, Prentice Hall Series in Artificial Intelligence (Boston Columbus Indianapolis: Pearson, 2016).

[60] Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (Viking, 2019): 11.

[61] Bernhard Rieder, "Scrutinizing an Algorithmic Technique: The Bayes Classifier as Interested Reading of Reality," *Information, Communication & Society* 20, no. 1 (January 2, 2017): 104.

[62] Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press, 2021): 9.

[63] Rita Raley and Jennifer Rhee, "Critical AI: A Field in Formation," *American Literature* 95, no. 2 (June 1, 2023): 188.

[64] Matteo Pasquinelli, "Three Thousand Years of Algorithmic Rituals: The Emergence of AI from the Computation of Space," *E-Flux*, no. 101 (June 2019).

[65] Crawford, *Atlas of AI* […]: 10.

[66] Ibid.

[67] Mackenzie, *Machine Learners* […]: 6.

[68] Matteo Pasquinelli and Vladan Joler, "The Nooscope Manifested: AI as Instrument of Knowledge Extractivism," *AI & SOCIETY* 36, no. 4 (December 2021): 1265.

[69] Mackenzie, *Machine Learners* […]: 55.

[70] Anika Shrivastava, Renu Rameshan, and Samar Agnihotri, "Latent Space Characterization of Autoencoder Variants" (arXiv, January 16, 2025): 1.

[71] Diederik P. Kingma and Max Welling, "Auto-Encoding Variational Bayes" (arXiv, December 10, 2022); Tero Karras, Samuli Laine, and Timo Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks" (arXiv, March 29, 2019).

[72] Jonathan Ho, Ajay Jain, and Pieter Abbeel, "Denoising Diffusion Probabilistic Models" (arXiv, December 16, 2020).

[73] N. Katherine Hayles, "Subversion of the Human Aura: A Crisis in Representation," *American Literature* 95, no. 2 (June 1, 2023): 258.

[74] Justin Joque, *Revolutionary Mathematics: Artificial Intelligence, Statistics and the Logic of Capitalism* (Verso Books, 2022): 134.

[75] Karin Knorr Cetina, *Epistemic Cultures: How the Sciences Make Knowledge* (Cambridge: Harvard University Press, 1999): 3.

[76] Crawford, *Atlas of AI* […]: 17.

[77] Lucy A. Suchman, "Epilogue" in *Human-Machine Reconfigurations: Plans and Situated Actions*, 2nd ed. (New York: Cambridge University Press, 2007); Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World*, Artificial Unintelligence (MIT Press Ltd, 2018); Wendy Hui Kyong Chun and Alex H. Barnett, *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition* (Cambridge, Massachusetts London, England: The MIT Press, 2021): 50-74

[78] Ted Chiang, "ChatGPT Is a Blurry JPEG of the Web," *The New Yorker*, February 9, 2023, https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web.

[79] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning* (Cambridge, Mass: The MIT Press, 2016): 110.

[80] Fazi, "Beyond Human […]": 10.

[81] Ibid., 11.

[82] Louise Amoore et al., "A World Model: On the Political Logics of Generative AI," *Political Geography* 113 (August 1, 2024): 4-5.

[83] Ibid., 7.

[84] Crawford, *Atlas of AI* […]: 221.

[85] Gillespie, "The Relevance of Algorithms,": 168.

[86] Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford University Press, 2007).

[87] Jackie Kay, Atoosa Kasirzadeh, and Shakir Mohamed, "Epistemic Injustice in Generative AI," *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 7 (October 16, 2024): 685.

[88] Joque, *Revolutionary Mathematics* […]: 134.

[89] Crawford, *Atlas of AI* […]: 11.

[90] Roland Meyer, "It's a Flat World. The Synthetic Realities of Sora," application/pdf, *Rrrreflect. Journal of Integrated Design Research; Special Issue 1*, 2024: 2.

[91] Mikael Brunila, "Taking AI into the Tunnels," e-flux, February 2025, [emphasis original].

[92] Crawford, *Atlas of AI* […]: 11.

[93] Hannes Bajohr, "Whoever Controls Language Models Controls Politics," *Hannes Bajohr* (blog), April 8, 2023.

[94] Walter Lippmann, *Public Opinion* (Routledge, 1997): 15.

[95] Ibid., 16.

[96] Amoore et al., "A World Model […]": 2; Joque, *Revolutionary Mathematics* […]: 144.

[97] Bernhard Rieder, "Scrutinizing an Algorithmic Technique: The Bayes Classifier as Interested Reading of Reality," *Information, Communication & Society* 20, no. 1 (January 2, 2017): 110.

[98] Matteo Pasquinelli, *The Eye of the Master: A Social History of Artificial Intelligence* (London: Verso Books, 2023).

[99] Leif Weatherby, "ChatGPT Is an Ideology Machine," Jacobin, April 17, 2023.

[100] Ibid.

[101] Mirzoeff, *The Right to Look* […]: 2.

[102] Ibid., 17.

[103] Ibid., 20.

[104] Louise Amoore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others* (Durham: Duke University Press, 2020): 17.

[105] Adrian MacKenzie and Anna Munster, "Platform Seeing: Image Ensembles and Their Invisualities," *Theory, Culture & Society* 36, no. 5 (September 1, 2019): 4.

[106] Amoore et al., "A World Model […]": 8.

[108] Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21 (New York, NY, USA: Association for Computing Machinery, 2021), 610–23.

[109] Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences*, Inside Technology (MIT Press, 1999): 1.

[110] Ibid., 5 & 319.

[111] Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan, "Semantics Derived Automatically from Language Corpora Contain Human-like Biases," *Science* 356, no. 6334 (April 14, 2017): 2.

[112] Pasquinelli, *The Eye of the Master* […]: 30.

[113] Batya Friedman and Helen Nissenbaum, "Bias in Computer Systems," *ACM Trans. Inf. Syst.* 14, no. 3 (July 1, 1996): 336.

[114] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Publishing Group, 2018).

[115] Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (Conference on Fairness, Accountability and Transparency, PMLR, 2018), 77–91.; Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press, 2018).

[116] Ruha Benjamin, *Race after Technology: Abolitionist Tools for the New Jim Code* (Cambridge: Polity Press, 2019): 6.

[117] Wendy Hui Kyong Chun and Alex H. Barnett, *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition* (Cambridge: The MIT Press, 2021): 24.

[118] Niharika Jain et al., "Imagining an Engineer: On GAN-Based Data Augmentation Perpetuating Biases" (arXiv, November 9, 2018); Sachit Menon et al., "PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models" (arXiv, July 20, 2020): 8; Vongani H. Maluleke et al., "Studying Bias in GANs through the Lens of Race" (arXiv, September 14, 2022).

[119] Alexandra Sasha Luccioni et al., "Stable Bias: Analyzing Societal Representations in Diffusion Models" (arXiv, November 9, 2023); Sahil Kuchlous, Marvin Li, and Jeffrey G. Wang, "Bias Begets Bias: The Impact of Biased Embeddings on Diffusion Models" (arXiv, September 15, 2024), https://doi.org/10.48550/arXiv.2409.09569.

[120] Mohammad Nadeem et al., "Gender Bias in Text-to-Video Generation Models: A Case Study of Sora" (arXiv, January 10, 2025).

[121] Bender et al., "On the Dangers of Stochastic Parrots […]": 613.; Christine Basta, Marta R. Costa-jussà, and Noe Casas, "Evaluating the Underlying Gender Bias in Contextualized Word Embeddings" (arXiv, April 18, 2019).

[122] Deepak Kumar, Yousef AbuHashem, and Zakir Durumeric, "Watch Your Language: Investigating Content Moderation with Large Language Models" (arXiv, January 17, 2024); Luca Rettenberger, Markus Reischl, and Mark Schutera, "Assessing Political Bias in Large Language Models" (arXiv, June 5, 2024); Shangbin Feng et al., "From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ed. Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (ACL 2023, Toronto, Canada: Association for Computational Linguistics, 2023)

[123] Nico Grant, "Google Chatbot's A.I. Images Put People of Color in Nazi-Era Uniforms," *The New York Times*, February 22, 2024, sec. Technology.

[124] "Google to Fix AI Picture Bot after 'woke' Criticism," BBC, web. February 21, 2024.

[125] Shangbin Feng et al., "From Pretraining Data to Language Models to Downstream Tasks […]: 11742.

[126] David Rozado, "Danger in the Machine: The Perils of Political and Demographic Biases Embedded in AI Systems," Issue Brief (New York: The Manhattan Institute, March 2023).

[127] Jochen Hartmann, Jasper Schwenzow, and Maximilian Witte, "The Political Ideology of Conversational AI: Converging Evidence on ChatGPT's pro-Environmental, Left-Libertarian Orientation" (arXiv, January 5, 2023); Yifei Liu, Yuang Panwang, and Chao Gu, "'Turning Right'? An Experimental Study on the Political Value Shift in Large Language Models," *Humanities and Social Sciences Communications* 12, no. 1 (February 10, 2025): 1–10.

[128] Bender et al., "On the Dangers of Stochastic Parrots […]": 613-614.

[129] Kate Crawford and Trevor Paglen, "Excavating AI: The Politics of Images in Machine Learning Training Sets," *AI & SOCIETY* 36, no. 4 (December 1, 2021): 1105–16.

[130] Milagros Miceli and Julian Posada, "The Data-Production Dispositif" (arXiv, May 24, 2022)

[131] Ibid., 1113.

[132] Amoore et al., "A World Model […]": 3.

[133] Ibid., 5; Valentin Hofmann et al., "Dialect Prejudice Predicts AI Decisions about People's Character, Employability, and Criminality" (arXiv, March 1, 2024).

[134] Chun, *Discriminating Data* […]: 57.

[135] Amoore et al., "A World Model […]": 4.

[136] Fabian Offert and Thao Phan, "A Sign That Spells: DALL-E 2, Invisual Images and The Racial Politics of Feature Space" (arXiv, October 26, 2022): 3.

[137] George Lipsitz, qtd. in Yarden Katz, *Artificial Whiteness: Politics and Ideology in Artificial Intelligence* (New York: Columbia University Press, 2020): 24.

[138] Crawford, *Atlas of AI* […]: 131.

[139] Bajohr, "Whoever Controls Language Models Controls Politics,"

[140] Kasy, "The Political Economy of AI […]": 12.

[141] Meredith Whittaker, "The Steep Cost of Capture," SSRN Scholarly Paper (Rochester, NY, 2021).

[142] Ben Green, "The Flaws of Policies Requiring Human Oversight of Government Algorithms," *Computer Law & Security Review* 45 (July 1, 2022).

[143] Jacques Rancière, *The Politics of Aesthetics: The Distribution of the Sensible* (Continuum, 2004): 12.

[144] Mirzoeff, *The Right to Look* […]: 23.

[145] Pasquinelli, *The Eye of the Master* […]: 17, 23.

[146] Alon Halevy, Peter Norvig, and Fernando Pereira, "The Unreasonable Effectiveness of Data," *IEEE Intelligent Systems* 24, no. 2 (March 2009): 9.

[147] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, vol. 25 (Curran Associates, Inc., 2012).

[148] Ibid., 2.

[149] danah boyd and Kate Crawford, "Critical Questions for Big Data," Information, Communication & Society 15, no. 5 (June 2012): 662–79

[150] Gaël Varoquaux, Alexandra Sasha Luccioni, and Meredith Whittaker, "Hype, Sustainability, and the Price of the Bigger-Is-Better Paradigm in AI" (arXiv, March 1, 2025).

[151] Elizabeth Gibney, "China's Cheap, Open AI Model DeepSeek Thrills Scientists," *Nature* 638, no. 8049 (January 23, 2025): 13–14, https://doi.org/10.1038/d41586-025-00229-6.

[152] Yuanzhi Li et al., "Textbooks Are All You Need II: Phi-1.5 Technical Report" (arXiv, September 11, 2023), https://doi.org/10.48550/arXiv.2309.05463.

[153] Chen Sun et al., "Revisiting Unreasonable Effectiveness of Data in Deep Learning Era" (arXiv, August 4, 2017), https://doi.org/10.48550/arXiv.1707.02968.

[154] Maximilian Schreiner, "GPT-4 Architecture, Datasets, Costs and More Leaked," THE DECODER, July 11, 2023, https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked/.

[155] Doug Laney, "3D Data Management: Controlling Data Volume, Velocity and Variety," Research Note, Application Delivery Strategies (META Group Inc., February 6, 2001).

[156] Zuboff, *The Age of Surveillance Capitalism* […], (2017).

[157] Nick Srnicek, *Platform Capitalism* (John Wiley & Sons, 2017).

[158] Sarah Myers West, "Data Capitalism: Redefining the Logics of Surveillance and Privacy," *Business & Society* 58, no. 1 (January 1, 2019): 20–41.

[159] McKenzie Wark, *Capital Is Dead: Is This Something Worse?* (Verso Books, 2019).

[160] Ibid., 11.

[161] Sy Taffel, "Data and Oil: Metaphor, Materiality and Metabolic Rifts," *New Media & Society* 25, no. 5 (May 1, 2023): 980–98.

[162] See: Garfield Benjamin, "What We Do with Data: A Performative Critique of Data 'Collection,'" *Internet Policy Review* 10, no. 4 (2021): 1–27; Jathan Sadowski, "When Data Is Capital: Datafication, Accumulation, and Extraction," *Big Data & Society* 6, no. 1 (January 1, 2019); Ulises Ali Mejias and Nick Couldry, *Data Grab: The New Colonialism of Big Tech and How to Fight Back* (Chicago, IL: The University of Chicago Press, 2024).

[163] Trevor Paglen, "Invisible Images (Your Pictures Are Looking at You)," *The New Inquiry* (blog), December 8, 2016.

[164] West, "Data Capitalism […]": 20.

[165] Amandalynne Paullada et al., "Data and Its (Dis)Contents: A Survey of Dataset Development and Use in Machine Learning Research" (arXiv, December 9, 2020).

[166] Christopher Zirpoli, "Generative Artificial Intelligence and Copyright Law," *Copyright, Fair Use, Scholarly Communication, Etc.*, February 24, 2023; Mark A. Lemley and Bryan Casey, "Fair Learning," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, January 30, 2020); Mehtab Khan and Alex Hanna, "The Subjects and Stages of AI Dataset Development: A Framework for Dataset Accountability," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, September 13, 2022).

[167] Melany Amarikwa, "Internet Openness at Risk: Generative AI's Impact on Data Scraping," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, February 5, 2024).

[168] Dulong de Rosnay (Mélanie) and Stalder (Felix), "Digital Commons," *Internet Policy Review* 9, no. 4 (December 17, 2020); Frédérique Dubé, "The Paradox of the Commons? - Part 1," PRAXIS Research Note, February 13, 2025; Saffron Huang and Divya Siddarth, "Generative AI and the Digital Commons" (arXiv, March 20, 2023).

[169] Pasquinelli, *The Eye of the Master* […]: 16.

[170] Ibid., 17.

[171] Jess Weatherbed, "Meta Fed Its AI on Almost Everything You've Posted Publicly since 2007," The Verge, September 12, 2024.

[172] Mary L. Gray and Siddharth Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* (Harper Collins, 2019); Paola Tubaro, Antonio A Casilli, and Marion Coville, "The Trainer, the Verifier, the Imitator: Three Ways in Which Human Platform Workers Support Artificial Intelligence," *Big Data & Society* 7, no. 1 (January 1, 2020)

[173] Karl Marx, "Estranged Labor," in *Economic and Philosophic Manuscripts of 1844*, trans. Martin Milligan (Progress Publishers, 1959)

[174] Alan Mathison Turing, "Lecture on the Automatic Computing Engine," in *The Essential Turing* (Oxford University Press, 2004): 392.

[175] David M. Berry and James Stockman, "Schumacher in the Age of Generative AI: Towards a New Critique of Technology," *European Journal of Social Theory*, March 1, 2024: 13.

[176] Neda Atanasoski and Kalindi Vora, *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures* (Duke University Press, 2019): 5.

[177] Ibid., 6.

[178] Thomas Carlyle, *On Heroes, Hero Worship, and the Heroic in History* (Yale University Press, 2013).

[179] Mirzoeff, *The Right to Look* […]: 4. [emphasis added]

[180] See: Hanna L. Grønneberg and Ana Alacovska, "Critical Dataset and Machine Learning Art," *Morals & Machines* 2, no. 2 (2022): 22–31; Kate Crawford and Trevor Paglen, "Excavating AI: The Politics of Images in Machine Learning Training Sets," *AI & SOCIETY* 36, no. 4 (December 1, 2021): 1105–16; Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton, "Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development," *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW2 (October 13, 2021): 1–37.

[181] Amanda Meng and Carl DiSalvo, "Grassroots Resource Mobilization through Counter-Data Action," *Big Data & Society* 5, no. 2 (July 1, 2018); "About - Data For Black Lives," D4BL, accessed March 14, 2025, https://d4bl.org/about. ; "About Us – The Markup," accessed March 14, 2025, https://themarkup.org/about.

[182] Matt Burgess, "How to Stop Your Data From Being Used to Train AI," *Wired*, accessed March 14, 2025, https://www.wired.com/story/how-to-stop-your-data-from-being-used-to-train-ai/; Shawn Shan et al., "Nightshade: Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models" (arXiv, April 29, 2024).

[183] Alex Pentland et al., "Data Cooperatives: Digital Empowerment of Citizens and Workers," Whitepaper, MIT Connection Science (Massachusetts Institute of Technology, January 2, 2019)

[184] Mirzoeff, *The Right to Look* […]: 66.

[185] Mackenzie, *Machine Learners* […]: 17.

[186] Michel Foucault, *Discipline and Punish: The Birth of the Prison* (New York: Vintage Books, 1995).

[187] William S. Burroughs, *The Electronic Revolution*, (Cambridge: Blackmoor Head Press, 1971)

[188] Gilles Deleuze, "Postscript on the Societies of Control," *October* 59 (1992): 4-5.

[189] Ulrich Beck, *Risk Society: Towards a New Modernity* (SAGE, 1992).

[190] Kevin D. Haggerty and Richard V. Ericson, "The Surveillant Assemblage," *The British Journal of Sociology* 51, no. 4 (2000): 605–22.

[191] Ibid., 613.

[192] Mackenzie, *Machine Learners* […]: 17.

[193] Mark Andrejevic, *Automated Media* (London ; New York, NY: Routledge, 2020): 74.

[194] Louise Amoore, "Lines of Sight: On the Visualization of Unknown Futures," *Citizenship Studies* 13, no. 1 (February 1, 2009): 22.

[195] Ibid.

[196] Antoinette Rouvroy, Thomas Berns, and Liz Carey-Libbrecht, "Algorithmic governmentality and prospects of emancipation:Disparateness as a precondition for individuation through relationships?," *Réseaux* 177, no. 1 (2013): 163–96.

[197] Benjamin, *Race After Technology* […]; Eubanks, *Automating Inequality* […]; Meredith Broussard, *More than a Glitch: Confronting Race, Gender, and Ability Bias in Tech* (MIT Press, 2023).

[198] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami, "Mining Association Rules between Sets of Items in Large Databases," in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data* (SIGMOD, Washington D.C. USA: ACM, 1993), 207–16.

[199] Zuboff, *The Age of Surveillance Capitalism* […]: 353, 354.

[200] Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You*, (Penguin Press, 2011); Taina Bucher, *If...Then: Algorithmic Power and Politics*, Oxford Studies in Digital Politics (Oxford University Press, USA, 2018).

[201] Robert Gorwa, Reuben Binns, and Christian Katzenbach, "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance," *Big Data & Society* 7, no. 1 (January 1, 2020); Tarleton Gillespie, "Content Moderation, AI, and the Question of Scale," *Big Data & Society* 7, no. 2 (July 1, 2020).

[202] Karen Yeung, "'Hypernudge': Big Data as a Mode of Regulation by Design," *Information, Communication & Society* 20, no. 1 (January 2, 2017): 118–36; Richard H. Thaler and Cass R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Yale University Press, 2008).

[203] Floridi Luciano, "Hypersuasion – On AI's Persuasive Power and How to Deal with It," *Philosophy & Technology* 37, no. 2 (May 17, 2024): 64.

[204] Zuboff, *The Age of Surveillance Capitalism* […]: 362. See also: Meredith Whittaker, "The Steep Cost of Capture," SSRN Scholarly Paper (Rochester, NY, 2021).

[205] Sarah Brayne, *Predict and Surveil: Data, Discretion, and the Future of Policing* (Oxford University Press, 2021).

[206] Kathleen McKendrick, "Artificial Intelligence: Prediction and Counterterrorism," Research Report (London: The Royal Institute of International Affairs, August 2019).

[207] Matthew Cole, Hugo Radice, and Charles Umney, "The Political Economy of Datafication and Work: A New Digital Taylorism?," *Socialist Register* 57 (2021).

[208] Valentin Weber, "Data-Centric Authoritarianism: How China's Development of Frontier Technologies Could Globalize Repression" (Washington, D.C: National Endowment for Democracy, February 11, 2025).

[209] Rui Hou and Diana Fu, "Sorting Citizens: Governing via China's Social Credit System," *Governance* 37, no. 1 (2024): 59–78.

[210] Niva Elkin-Koren and Maayan Perel, "Separation of Functions for AI: Restraining Speech Regulation by Online Platforms," *Lewis & Clark Law Review* 24 (2020): 865.

[211] Ronald J. Deibert, "The Road to Digital Unfreedom: Three Painful Truths About Social Media," *Journal of Democracy* 30, no. 1 (2019): 25.

[212] Simeon Gilding, "De-Risking Authoritarian AI," Policy Brief (Australian Strategic Policy Institute, July 2023).

[213] Winner*, Do Artifacts Have Politics?*: 129.

[214] Ibid., 130.

[215] Ibid., 123.

[216] Joshua Tucker et al., "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature," *SSRN Electronic Journal*, 2018: 4.

[217] Rachel Kuo and Alice Marwick, "Critical Disinformation Studies: History, Power, and Politics," *Harvard Kennedy School Misinformation Review*, August 12, 2021.

[218] Jennifer Kavanagh and Michael D. Rich, "Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life" (RAND Corporation, January 15, 2018).

[219] Adrienne Thompson, "The Existential Threat of AI-Enhanced Disinformation Operations," *Center for Security and Emerging Technology* (blog), July 11, 2022.

[220] Renee DiResta, "The Digital Maginot Line," *RibbonFarm* (blog), November 28, 2018.

[221] Adrien Friggeri et al., "Rumor Cascades," *Proceedings of the International AAAI Conference on Web and Social Media* 8, no. 1 (May 16, 2014): 101–10; Killian L. McLoughlin and William J. Brady, "Human-Algorithm Interactions Help Explain the Spread of Misinformation," *Current Opinion in Psychology* 56 (April 1, 2024); Soroush Vosoughi, Deb Roy, and Sinan Aral, "The Spread of True and False News Online," *Science* 359, no. 6380 (March 9, 2018): 1146–51;

[222] Adrian Chen, "The Agency," *The New York Times Magazine*, June 2, 2015; Carole Cadwalladr and Emma Graham-Harrison, "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach," *The Guardian*, March 17, 2018, sec. News.; Marco T. Bastos and Dan Mercea, "The Brexit Botnet and User-Generated Hyperpartisan News," *Social Science Computer Review* 37, no. 1 (February 1, 2019): 38–54.

[223] John Akers et al., "Technology-Enabled Disinformation: Summary, Lessons, and Recommendations" (arXiv, January 3, 2019); Noémi Bontridder and Yves Poullet, "The Role of Artificial Intelligence in Disinformation," *Data & Policy* 3 (January 2021).

[224] Orsolya Reich Calabrese Sofia, "Beyond Disinformation: How DSA Risk Assessments Ignore Democracy's Real Threats," Policy Analysis (Tech Policy Press, March 19, 2025).

[225] Robert Chesney and Danielle Keats Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," *SSRN Electronic Journal*, 2018: 1776.

[226] Don Fallis, "The Epistemic Threat of Deepfakes," *Philosophy & Technology* 34, no. 4 (December 1, 2021): 623–43.

[227] Cristian Vaccari and Andrew Chadwick, "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News," *Social Media + Society* 6, no. 1 (January 1, 2020): 1.

[228] Regina Rini, "Deepfakes and the Epistemic Backstop," *Philosopher's Imprint* 20, no. 24 (2020): 8.

[229] Megan Hyska, "The Politics of Past and Future: Synthetic Media, Showing, and Telling: The Politics of Past and Future: Synthetic Media...: M. Hyska," *Philosophical Studies* 182, no. 1 (January 2025): 150.

[230] Michael Hameleers, "Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting," *International Journal of Public Opinion Research* 36, no. 1 (March 1, 2024): 7 ;

[231] Recent studies include: Dipto Barman, Ziyi Guo, and Owen Conlan, "The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination," *Machine Learning with Applications* 16 (June 1, 2024): 10; Ivan Vykopal et al., "Disinformation Capabilities of Large Language Models," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2024, 14830–47.

[232] Thomas Zerback, Florian Töpfl, and Maria Knöpfle, "The Disconcerting Potential of Online Disinformation: Persuasive Effects of Astroturfing Comments and Three Strategies for Inoculation against Them," *New Media & Society* 23, no. 5 (May 1, 2021): 1080–98.

[233] Michal Klincewicz, Mark Alfano, and Amir Fard, "Slopaganda: The Interaction Between Propaganda and Generative Ai," *Filosofiska Notiser* 12, no. 1 (2025): 135–62.

[234] Nina Schick, *Deepfakes: The Coming Infocalypse* (Grand Central Publishing, 2020).

[235] Hal Berghel, "Generative Artificial Intelligence, Semantic Entropy, and the Big Sort," *Computer* 57, no. 1 (January 2024): 130–35; Kyree Leary, "An AI That Makes Fake Videos May Facilitate the End of Reality as We Know It," Futurism, December 8, 2017.

[236] Johan Farkas and Jannick Schou, *Post-Truth, Fake News and Democracy: Mapping the Politics of Falsehood*, Second edition (New York, NY: Routledge, 2024): 131.

[237] Andrea Bellemare and Jason Ho, "Social Media Firms Catching More Misinformation, but Critics Say They Could Be Doing More," *CBC News*, April 20, 2020. Monika Bickert, "Our Approach to Labeling AI-Generated Content and Manipulated Media," *Meta* (blog), April 5, 2024; Sarah A. Fisher, "Something AI Should Tell You – The Case for Labelling Synthetic Content," *Journal of Applied Philosophy* 42, no. 1 (2025): 272–86.

[238] Nandor Kiss, "The Twenty-Six Words That Created the Internet… and Then Maybe, Kind Of, Destroyed Society: Understanding and Reforming Section 230 of the Communications Decency Act," *Michigan Technology Law Review* 31, no. 1 (January 1, 2024): 131–65.

[239] Dogus Karabulut, Cagri Ozcinar, and Gholamreza Anbarjafari, "Automatic Content Moderation on Social Media," *Multimedia Tools and Applications* 82, no. 3 (January 1, 2023); Robert Gorwa, Reuben Binns, and Christian Katzenbach, "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance," *Big Data & Society* 7, no. 1 (January 1, 2020).

[240] Gabriel Nicholas, "Shedding Light on Shadowbanning" (Washington, D.C: Center for Democracy & Technology, May 19, 2022).

[241] David Ghiurău and Daniela Elena Popescu, "Distinguishing Reality from AI: Approaches for Detecting Synthetic Content," *Computers* 14, no. 1 (January 2025); Firoj Alam et al., "A Survey on Multimodal Disinformation Detection" (arXiv, September 28, 2022);

[242] Farkas and Schou, *Post-Truth, Fake News and Democracy*: 139.

[243] Meta, "Partnering with Third-Party Fact-Checkers," *Meta Journalism Project* (blog), March 23, 2020.

[244] Gordon Pennycook and David G. Rand, "Accuracy Prompts Are a Replicable and Generalizable Approach for Reducing the Spread of Misinformation," *Nature Communications* 13, no. 1 (April 28, 2022).

[245] Robert McPhedran et al., "Psychological Inoculation Protects against the Social Media Infodemic," *Scientific Reports* 13, no. 1 (April 8, 2023): 5780.

[246] Johannes Sedlmeir et al., "Battling Disinformation with Cryptography," *Nature Machine Intelligence* 5, no. 10 (October 2023): 1056–57.

[247] Reza Montasari, "The Dual Role of Artificial Intelligence in Online Disinformation: A Critical Analysis," in *Cyberspace, Cyberterrorism and the International Security in the Fourth Industrial Revolution: Threats, Assessment and Responses*, ed. Reza Montasari (Cham: Springer International Publishing, 2024), 229–40.

[248] Rachel Kuo and Alice Marwick, "Critical Disinformation Studies: History, Power, and Politics," *Harvard Kennedy School Misinformation Review*, August 12, 2021; Robert Mejia, Kay Beckermann, and Curtis Sullivan, "White Lies: A Racial History of the (Post)Truth," *Communication and Critical/Cultural Studies* 15, no. 2 (April 3, 2018): 109–26.

[249] W. Lance Bennett and Steven Livingston, "A Brief History of the Disinformation Age: Information Wars and the Decline of Institutional Authority," in *The Disinformation Age*, ed. Steven Livingston and W. Lance Bennett, SSRC Anxieties of Democracy (Cambridge: Cambridge University Press, 2020): 4.

[250] Ibid., 8.

[251] Ibid., 19.

[252] Jacques Ellul described 'modern propaganda' (i.e. public relations) as a "modern scientific system," wherein measures are quantitatively assessed and adjusted to achieve maximum effect. Jacques Ellul, *Propaganda: The Formation of Men's Attitudes* (Knopf Doubleday Publishing Group, 1965): 4.

[253] Bennett and Livingston, "A Brief History of the Disinformation Age […]": 14.

[254] Herman and Chomsky, *Manufacturing Consent* […].

[255] Kuo and Marwick, "Critical Disinformation Studies […]": 4.

[256] Bennett and Livingston, "A Brief History of the Disinformation Age […]": 20.

[257] Sheldon S. Wolin, *Democracy Incorporated: Managed Democracy and the Specter of Inverted Totalitarianism - New Edition* (Princeton University Press, 2017).

[259] Bennett and Livingston, "A Brief History of the Disinformation Age […]": 11.

[260] Gallup Inc, "Confidence in Institutions," Gallup.com, 2024, https://news.gallup.com/poll/1597/Confidence-Institutions.aspx.

[261] Peter Dahlgren, "Media, Knowledge and Trust: The Deepening Epistemic Crisis of Democracy," *Javnost - The Public* 25, no. 1–2 (April 3, 2018): 24.

[262] Svenja Hense and Armin Schäfer, "Unequal Representation and the Right-Wing Populist Vote," in *Contested Representation: Challenges, Shortcomings and Reforms*, ed. Claudia Landwehr, Thomas Saalfeld, and Armin Schäfer (Cambridge University Press, 2022)

[263] Tim Hayward, "The Problem of Disinformation: A Critical Approach," *Social Epistemology* 0, no. 0 (2024): 19.

[264] Joseph Bernstein, "Bad News: Selling the Story of Disinformation," *Harper's Magazine*, accessed March 18, 2024, https://harpers.org/archive/2021/09/bad-news-selling-the-story-of-disinformation/.

[265] J. M. Berger, "Our Consensus Reality Has Shattered," *The Atlantic* (blog), October 9, 2020, https://www.theatlantic.com/ideas/archive/2020/10/year-living-uncertainly/616648/.

[266] Marc Tuters, "Fake News," *Krisis: Journal for Contemporary Philosophy* 38, no. 2 (2018): 60.

[267] Herman and Chomsky, *Manufacturing Consent* […].

[268] Jon Askonas, "What Happened to Consensus Reality?," *The New Atlantis*, Spring 2022 (June 30, 2022), https://www.thenewatlantis.com/publications/what-happened-to-consensus-reality."

[269] Michel Foucault, *The Order of Things: An Archaeology of the Human Sciences* (Psychology Press, 2002).

[270] Jürgen Habermas, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*, trans. Thomas Burger (Cambridge, MA: MIT Press, 1989): 27.

[271] Andreas Jungherr and Ralph Schroeder, "Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy," *Social Media + Society* 7, no. 1 (January 1, 2021): 4.

[272] Jungherr and Schroeder, "Disinformation and the Structural […],": 4.

[273] A. J. Bauer and Anthony Nadler, "What Is Disinformation to Democracy?," *Center for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill*, January 23, 2023, https://citap.pubpub.org/pub/u4qzry5p/release/1.

[274] Jungherr and Schroeder, "Disinformation and the Structural […],": 5.

[275] Théophile Lenoir and Chris Anderson, "Introduction Essay: What Comes After Disinformation Studies," *Center for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill*, January 23, 2023: 4. https://citap.pubpub.org/pub/oijfl3sv/release/1.

[276] Bernstein, "Bad News […],"; Jack Bratich, "Civil Society Must Be Defended: Misinformation, Moral Panics, and Wars of Restoration," *Communication, Culture and Critique* 13, no. 3 (September 1, 2020): 311–32.

[277] Johan Farkas and Jannick Schou, "Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood," *Javnost - The Public* 25, no. 3 (July 3, 2018): 298.

[278] Kuo and Marwick, "Critical Disinformation Studies […]": 2.

[279] Bruce Mutsvairo et al., "'Factforward:' Foretelling the Future of Africa's Information Disorder," *Center for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill*, January 23, 2023, https://citap.pubpub.org/pub/lopohgbv/release/1.; Grisel Salazar, "Rethinking Disinformation for the Global South: Towards a Particular Research Agenda," in *State-Sponsored Disinformation Around the Globe*, ed. Martin Echeverria, Sara Santamaria, and Daniel Hallin (Routledge, 2024).

[280] See: Amy Orben, "The Sisyphean Cycle of Technology Panics," *Perspectives on Psychological Science* 15, no. 5 (September 1, 2020): 1143–57; Felix M Simon and Chico Q Camargo, "Autopsy of a Metaphor: The Origins, Use and Blind Spots of the 'Infodemic,'" *New Media & Society* 25, no. 8 (August 1, 2023): 2219–40; Jungherr and Schroeder, "Disinformation and the Structural […],"; among others.

[281] Stanley Cohen, *Folk Devils and Moral Panics* (New York: Routledge, 1972)

[282] Stuart Hall et al., *Policing the Crisis: Mugging, the State, and Law and Order* (The Macmillan Press, 1978)

[283] Cohen, *Folk Devils and Moral Panics*: 2.

[284] Marita Sturken, Douglas Thomas, and Sandra Ball-Rokeach, *Technological Visions: The Hopes and Fears That Shape New Technologies* (Temple University Press, 2004): 1.

[285] Felix M. Simon, Sacha Altay, and Hugo Mercier, "Misinformation Reloaded? Fears about the Impact of Generative AI on Misinformation Are Overblown," *Harvard Kennedy School Misinformation Review*, October 18, 2023; Joshua Habgood-Coote, "Deepfakes and the Epistemic Apocalypse," *Synthese* 201, no. 3 (March 9, 2023): 103; Sacha Altay, Manon Berriche, and Alberto Acerbi, "Misinformation on Misinformation: Conceptual and Methodological Challenges," *Social Media + Society* 9, no. 1 (January 1, 2023)

[286] Jennifer Allen et al., "Evaluating the Fake News Problem at the Scale of the Information Ecosystem," *Science Advances* 6, no. 14 (April 3, 2020): 3539; Sacha Altay, Rasmus Kleis Nielsen, and Richard Fletcher, "Quantifying the 'Infodemic': People Turned to Trustworthy News Outlets during the 2020 Coronavirus Pandemic," *Journal of Quantitative Description: Digital Media* 2 (August 26, 2022).

[287] Andrew Guess, Jonathan Nagler, and Joshua Tucker, "Less than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook," *Science Advances* 5, no. 1 (January 9, 2019): eaau4586; Nir Grinberg et al., "Fake News on Twitter during the 2016 U.S. Presidential Election," *Science* 363, no. 6425 (January 25, 2019): 374–78

[288] Alberto Acerbi, Sacha Altay, and Hugo Mercier, "Research Note: Fighting Misinformation or Fighting for Information?," *Harvard Kennedy School Misinformation Review*, January 12, 2022, https://doi.org/10.37016/mr-2020-87.

[289] Brendan Nyhan, "Facts and Myths about Misperceptions," *Journal of Economic Perspectives* 34, no. 3 (August 2020): 220–36; Gregory Eady et al., "Exposure to the Russian Internet Research Agency Foreign Influence Campaign on Twitter in the 2016 US Election and Its Relationship to Attitudes and Voting Behavior," Nature Communications 14, no. 1 (January 9, 2023); Levi Boxell, Matthew Gentzkow, and Jesse M. Shapiro, "A Note on Internet Use and the 2016 U.S. Presidential Election Outcome," PLOS ONE 13, no. 7 (July 18, 2018): e0199571.

[290] Maarten Hillebrandt, "The Communicative Model of Disinformation: A Literature Note," SSRN Scholarly Paper (Rochester, NY, March 5, 2021): 4.

[291] Tim Hayward, "The Problem of Disinformation: A Critical Approach," *Social Epistemology* 0, no. 0 (2024): 16.

[292] Anna Broinowski, "Deepfake Nightmares, Synthetic Dreams: A Review of Dystopian and Utopian Discourses Around Deepfakes, and Why the Collapse of Reality May Not Be Imminent—Yet," *Journal of Asia-Pacific Pop Culture* 7, no. 1 (2022): 109–39; Joshua Habgood-Coote, "Deepfakes and the Epistemic Apocalypse," *Synthese* 201, no. 3 (March 9, 2023).

[293] Simon et al., "Misinformation reloaded? […],": 3.

[294] Chesney and Citron, "Deep Fakes […],": 1785.

[295] Britt Paris and Joan Donovan, "Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence," *Data & Society*, September 2019; Mia Fineman, National Gallery of Art (U.S.), and Museum of Fine Arts Houston, *Faking It: Manipulated Photography Before Photoshop* (Metropolitan Museum of Art, 2012); Walter Scheirer, *A History of Fake Things on the Internet*, 1st ed. (Stanford University Press, 2023).

[296] Mathias Osmundsen et al., "Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter," *American Political Science Review* 115, no. 3 (August 2021): 999–1015.

[297] Habgood-Coote, "Deepfakes and the Epistemic Apocalypse,": 103.

[298] Jungherr and Schroeder, "Disinformation and the Structural […],": 5.

[299] Chin-Kuei Tsui, "The Securitization of Disinformation: Taiwan's Fake News Phenomenon and Anti-Disinformation Initiatives," *Tamkang Journal of International Affairs* 24, no. 2 (October 2020): 1–53; Gabrielle Lim, "SECURITIZE/ COUNTER-SECURITIZE" (New York: Data & Society, March 25, 2020).

[300] Barry Buzan, Ole Wæver, and Jaap de Wilde, *Security: A New Framework for Analysis* (Lynne Rienner Publishers, 2022): 21.

[301] Lim, "SECURITIZE/COUNTER-SECURITIZE,": 41.

[302] DiResta, "The Digital Maginot Line,".

[303] Clinton Castro, Pham, and Alan Rubel, "Epistemic Paternalism Online," in *Epistemic Paternalism: Conceptions, Justifications and Implications*, ed. Guy Axtell and Amiel Bernal (Rowman & Littlefield, 2020).

[304] Farkas and Schou, *Post-Truth, Fake News and Democracy*: 137; Montasari, "The Dual Role of Artificial Intelligence in Online Disinformation […],";

[305] Jordi Calvet-Bademunt and Jacob Mchangama, "Freedom of Expression in Generative AI: A Snapshot of Content Policies" (The Future of Free Speech, February 2024).

[306] Emma J Llansó, "No Amount of 'AI' in Content Moderation Will Solve Filtering's Prior-Restraint Problem," *Big Data & Society* 7, no. 1 (January 1, 2020): 3.

[307] Jamie Susskind, "What Discourse Regulation by Social Media Giants Means For Democratic Societies," *Literary Hub* (blog), July 5, 2022, https://lithub.com/what-discourse-regulation-by-social-media-giants-means-for-democratic-societies/.

[308] Danielle Allen and E. Glen Weyl, "The Real Dangers of Generative AI," *Journal of Democracy* 35, no. 1 (2024): 147–62; Konrad Bleyer-Simon, "Government Repression Disguised as Anti-Disinformation Action: Digital Journalists' Perception of COVID-19 Policies in Hungary," *Journal of Digital Media & Policy* 12, no. COVID-19 and Digital Media Policy (March 1, 2021): 159–76.

[309] Tarleton Gillespie, "Content Moderation, AI, and the Question of Scale," Big Data & Society 7, no. 2 (July 1, 2020): 4.

[310] Lenoir and Anderson, "Introduction Essay: What Comes After Disinformation Studies,": 7.

[311] Ibid., 8. [emphasis added]

[312] Bennett and Livingston, "A Brief History […],": 32.

[313] Jesse Hirsh, "128: The End of the Public?," Substack newsletter, *Metaviews: Future of Authority* (blog), March 12, 2025, https://metaviews.substack.com/p/the-end-of-the-public?utm_medium=web.

[314] Jacques Rancière, *Hatred of Democracy* (Verso Books, 2014): 80.

[315] Ibid.

[317] Oliver P. Richmond, "A Prelude to Revisionism? The Stalemated Peace Model and the Emergence of Multipolarity in International Order," *Contemporary Security Policy*, April 3, 2025: 201.

[318] Ibid., 215.

[319] Bernard E. Harcourt, *The Counterrevolution: How Our Government Went to War Against Its Own Citizens* (Basic Books, 2018); Kristian Williams, "Insurgency, counterinsurgency, and whatever comes next," in *Life during Wartime: Resisting Counterinsurgency*, ed. Kristian Williams, William Munger, and Lara Messersmith (Oakland, CA, USA: AK Press, 2013): 5-25.

[320] Jack Bratich, "Civil Society Must Be Defended: Misinformation, Moral Panics, and Wars of Restoration," *Communication, Culture and Critique* 13, no. 3 (September 1, 2020): 311–32; Alexei Abrahams and Gabrielle Lim, "Repress/Redress: What the 'War on Terror' Can Teach Us about Fighting Misinformation," *Harvard Kennedy School Misinformation Review*, July 22, 2020.

[321] Jayson Harsin, "Regimes of Posttruth, Postpolitics, and Attention Economies," *Communication, Culture & Critique* 8, no. 2 (2015): 327–33.

[322] Ibid., 332.

[323] Jungherr and Schroeder, "Disinformation and the Structural […],": 5.

[324] MacKenzie & Munster, "Platform Seeing […]."

[325] U.S. Senate. Committee on the Judiciary. *The Censorship Industrial Complex*. Hearing before the Subcommittee on the Constitution, 118th Cong., 2nd sess., March 25, 2025. https://www.judiciary.senate.gov/committee-activity/hearings/the-censorship-industrial-complex.

[326] Dan Williams, "There Is No 'Censorship Industrial Complex,'" *Conspicuous Cognition* (blog), July 28, 2024, https://www.conspicuouscognition.com/p/there-is-no-censorship-industrial.

[327] Jack Z. Bratich, *Conspiracy Panics: Political Rationality and Popular Culture* (Albany: SUNY Press, 2008): 16.

[328] Michel Foucault, *The History of Sexuality: An Introduction* (Pantheon Books, 1978): 93-95.

[329] Mirzoeff, *The Right to Look* […]: 3.

PART ONE

[1] Nicholas Mirzoeff. "Introduction The Right to Look, or, How to Think With and Against Visuality" in *The Right to Look: A Counterhistory of Visuality*. (Durham: Duke University Press, 2011), 6.

[2] Aristotle, *Nicomachean Ethics*, ed. Roger Crisp (Cambridge University Press, 2000)

[3] Ernst Cassirer, *An Essay on Man: An Introduction to a Philosophy of Human Culture* (New Haven, CT: Yale University Press, 1992), 24.

[5] Ibid., 25.

[6] Ibid.

[7] Plato, "Allegory of the Cave" in *The Republic* (London: Penguin Books, 1987)

[8] Ibid., 517a.

[9] Jean Baudrillard, "The Precession of Simulacra" in *Simulacra and Simulation (*Ann Arbor: University of Michigan Press, 1994), 6.

[10] Ibid., 8.

[11] Ibid., 4.

[12] Martin Heidegger, "The Age of The World Picture" in *The Question Concerning Technology and Other Essays* (New York: Harper Torchbooks, 1977), 115.

[13] Ibid., 126-127. [emphasis mine]

[14] Ibid., 150.

[15] Ibid., 141.

[16] Ibid., 130.

[18] Stuart Hall. "Representation & the Media" ed. Sanjay Talreja, et al. *Media Education Foundation Transcript*. (Northhampton MA: Media Education Foundation, 1997), 7.

[19] Ibid.

[21] Mirzoeff, *The Right to Look* […]: 1.

[22] Stuart Hall, *Representation: Cultural Representations* […], 6.

[24] Murray Jacob Edelman. "The Cardinal Political Role of Art," in *From Art to Politics: How Artistic Creations Shape Political Conceptions*. (Chicago: University of Chicago Press, 2003), 1.

[25] Ibid.

[26] Jacques Ellul. *Propaganda: The formation of men's attitudes.* (New York: Knopf, 1965), 11.

[27] Ibid., 139.

[28] Ibid., 140.

[29] Stuart Hall. "Representation & the Media," 10.

[30] Nicholas Mirzoeff, *The Right to Look*, 3.

[31] Ibid., xv.

[32] Ibid., 2.

[33] Ibid., 6.

[34] Thomas Carlyle et al., *On Heroes, Hero Worship, and the Heroic in History* (Yale University Press, 2013)

[35] Nicholas Mirzoeff, "On Visuality," *Journal of Visual Culture* 5 (April 1, 2006): 57.

[37] Nicholas Mirzoeff, *The Right to Look*, 1. [emphasis mine]

[38] Ibid., 4.

[39] Dipesh Chakrabarty, *Provincializing Europe: Postcolonial Thought and Historical Difference* (Princeton (N.J.): Princeton University Press, 2007), 63.

[40] Nicholas Mirzoeff, *The Right to Look*, 23-24.

[42] Ibid., 9.

[43] Michel Foucault, *Discipline and Punish: The Birth of the Prison* (New York: Vintage Books, 1995): 82. [emphasis added]

[44] Ibid., 89.

[46] Nicholas Mirzoeff, *The Right to Look*, 3.

[48] Ibid., 5.

[49] See: Alexander R. Galloway, *Protocol: How Control Exists after Decentralization*, Leonardo (Cambridge, Mass: MIT Press, 2004), 22; Gilles Deleuze, "Postscript on the Societies of Control," *October* 59 (1992): 6.

[50] Galloway, *Protocol* […]: 27.

[51] Nicholas Mirzoeff, *The Right to Look*, 56-57.

[52] Michel Foucault, *The Order of Things: An Archaeology of the Human Sciences* (New York: Pantheon Books, 1970).

[53] Nicholas Mirzoeff, *The Right to Look*, 65.

[54] Nicholas Mirzoeff, *The Right to Look*, 11.

[55] Ibid., 58.

[56] Ibid., 51.

[57] Ibid., 52.

[58] Ibid.

[59] Ibid. 79.

[60] Ibid., 61.

[61] Ibid., 76.

[62] Nicholas Mirzoeff, *The Right to Look*, 196.

[63] Ibid., 215.

[64] Lambert Adolphe Jacques Quetelet, *A Treatise on Man and the Development of His Faculties* (W. & R. Chambers, 1842).

[65] Frederick Douglass, *The Color Line* (The North American Review, 1881)

[66] W. E. B. Dubois, "Of the Dawn of Freedom" in *The Souls of Black Folk* (Simon and Schuster, 2005) 17-43.

[67] Nicholas Mirzoeff, *The Right to Look*, 216.

[68] Leslie Stroebel and Richard D. Zakia, *The Focal Encyclopedia of Photography* (Focal Press, 1993), 6.

[69] Scott Walden, "Objectivity in Photography," *The British Journal of Aesthetics* 45, no. 3 (July 1, 2005): 259.

[70] Susan Sontag. *On Photography*. (New York: Anchor Books, 1977), 154.

[71] Michel Foucault, *Discipline and Punish* […]

[72] Allan Sekula, “The Body and the Archive,” *October* 39 (1986): 7.

[73] Ibid., 18.

[74] Ibid., 10.

[75] John Tagg, *The Burden of Representation: Essays on Photographies and Histories* (U of Minnesota Press, 1993), 6.

[76] Michel Foucault, *Discipline and Punish*, 23.

[77] Michel Foucault, "Truth and Power" in *Power/Knowledge: Selected Interviews & Other Writings From 1972- 1977*, ed. Colin Gordon (New York: Pantheon Books, 1972): 131, 133.

[78] John Tagg, *The Burden of Representation*, 98.

[79] Ibid., 99-100. [emphasis mine]

[81] Ibid., 100.

[82] Ibid. [emphasis mine]

[83] Ibid., 189.

[85] Susan Sontag. *On Photography*, 14.

[86] John Tagg, *The Burden of Representation*, 11.

[87] Nicholas Mirzoeff, *The Right to Look*, 233.

[88] Ibid., 238.

[89] Ibid., 221. [emphasis original]

[90] Allan Sekula, “The Body and the Archive”: 56.

[91] Ibid.

[93] Ibid., 28.

[94] Ibid., 17.

[95] Mirzoeff, *The Right to Look* […]: 18.

[96] Nicholas Mirzoeff, *The Right to Look*, 297. [emphasis mine]

[97] Ibid., 279.

[98] Ibid., 278, 295.

[100] United States Dept. of the Army, *The U.S. Army/Marine Corps Counterinsurgency Field Manual* (Chicago: The University of Chicago Press, 2007), xlv.

[101] Ibid., 90.

[103] Nicholas Mirzoeff, *The Right to Look*, 300.

[105] Michel Foucault, *The History of Sexuality. Volume 1: An Introduction*, trans. R. Hurley (New York: Random House, 1978), 138.

[106] “A HISTORY OF THE HAMLET EVALUATION SYSTEM (HES) | CIA FOIA (Foia.Cia.Gov),” accessed August 31, 2023, https://www.cia.gov/readingroom/document/cia-rdp80r01720r000200120001-2.pdf

[107] Myra Gray, “Defending the US with Biometrics,” *Infosecurity Today* 6, no. 6 (2009): 24–25, https://doi.org/10.1016/S1742-6847(09)70015-1.

[108] Nicholas Mirzoeff, *The Right to Look*, 282.

[109] William S. Burroughs, *The Electronic Revolution*, (Cambridge: Blackmoor Head Press, 1971)

[110] Gilles Deleuze, "Postscript on the Societies of Control," *October* 59 (1992): 4.

[111] Alex Williams, "Control Societies and Platform Logic.('Postscript on the Societies of Control' by Gilles Deleuze)," *New Formations*, no. 84–85 (2015): 215.

[112] Gilles Deleuze, "Postscript […]": 6.

[113] Edward L. Bernays, *Propaganda* (New York: H. Liveright, 1928); Jacques Ellul. *Propaganda: The Formation of Men's Attitudes* (New York: Knopf, 1965)

[114] Theodor W. Adorno and Max Horkheimer, "The Culture Industry: Enlightenment as Mass Deception," in *Dialectic of Enlightenment: Philosophical Fragments*, ed. Gunzelin Schmid Noerr, trans. Edmund Jephcott (Stanford: University of Stanford Press, 2002), 94–136.

[115] Guy Debord. *The Society of the Spectacle*. trans. Donald Nicholson-Smith. (NYC: Zone Books, 1994): 13.

[116] Jacques Ellul. *Propaganda* […]: 11.

[117] Ibid., 4.

[118] Ibid., 27.

[119] Alex Williams, "Control Societies and Platform Logic […]": 216.

[120] Alexander R. Galloway, *Protocol: How Control Exists after Decentralization* (Cambridge, Mass: MIT Press, 2004): 75. [emphasis mine]

[121] Ibid., 7, 3.

[122] Eugene Thacker, "Foreword: Protocol Is as Protocol Does" in *Protocol: How Control Exists after Decentralization*, by Alexander Galloway (Cambridge, Mass: MIT Press, 2004): xviii.

[123] Alexander R. Galloway, *Protocol: How Control Exists after Decentralization*: 9-10.

[125] The chapter "Physical Media" of *Protocol: How Control Exists after Decentralization* provides a thorough breakdown.

[126] Ibid., 52. [emphasis original]

[127] Ibid., 50. [emphasis original]

[128] Ibid., 95.

[129] Nick Srnicek, *Platform Capitalism* (Cambridge: Polity Press, 2017): 30.

[130] McKenzie Wark. "Capitalism—or Worse?" in *Capital Is Dead. Is This Something Worse?* (London: Verso, 2018): 41.

[131] Nick Srnicek, *Platform Capitalism*: 30.

[132] Alex Williams, "Control Societies and Platform Logic […]": 221.

[134] Ibid., 225.

[136] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," Nature 521, no. 7553 (May 2015): 436.

[137] Eugene Thacker, "Foreword: Protocol Is as Protocol Does" in *Protocol*: xvi.

[138] For example: Benjamin N Jacobsen, "Regimes of Recognition on Algorithmic Media," *New Media & Society* 25, no. 12 (December 1, 2023): 3641–56; Louise Amoore, "Neural Networks and Regimes of Recognition," in *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others* (Durham: Duke University Press, 2020); Taina Bucher, "Want to Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook," *New Media & Society* 14, no. 7 (2012): 1164–80.

[139] Safiya Umoja Noble. *Algorithms of Oppression: How Search Engines Reinforce Racism*. (New York: NYU Press, 2018).

[140] Ruha Benjamin. *Race after Technology: Abolitionist Tools for the New Jim Code*. (Cambridge: Polity Press, 2019): 6.

[141] Wendy Hui Kyong Chun. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition.* (Cambridge, MA:

MIT Press,

2021).

[142] Ruha Benjamin. *Race after Technology*: 10.

[143] Bernhard Rieder, "Scrutinizing an Algorithmic Technique: The Bayes Classifier as Interested Reading of Reality," *Information, Communication & Society* 20, no. 1 (January 2, 2017): 110.

[145] Ibid., 114.

[146] William Burroughs, "The Limits of Control." *Semiotext(e): Schizo-Culture* 3.2 (1978): 38.

[147] Ibid., 41.

[148] Nicholas Mirzoeff, *The Right to Look*, 285.

[150] Alexander R. Galloway, *Protocol: How Control Exists after Decentralization*: 75.

[151] Ibid., 76.

[152] Ibid., 64.

[153] Ulises A. Mejias, Nick Couldry, *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*, 1st ed., Culture and Economic Life (Redwood City: Stanford University Press, 2019): xiv.

[154] Nicholas Mirzoeff, *The Right to Look*: 1.

[155] Ibid., 4.

[156] Ibid.

[158] Ibid., 297.

[159] Adrian MacKenzie and Anna Munster, "Platform Seeing: Image Ensembles and Their Invisualities," *Theory, Culture & Society* 36, no. 5 (September 1, 2019): 3.

[160] Nicholas Mirzoeff, *The Right to Look*: 296.

[161] See: Shane Denson, *Discorrelated Images* (Durham, NC: Duke University Press, 2020).

[162] Ibid.

[163] Ibid., 280. [emphasis mine]

[164] Dominique Routhier, Lila Lee-Morrison, and Kathrin Maurer, "Automating Visuality: An Introduction," *MAST The Journal of Media Art Study and Theory* 3, no. 1 (April 2022): 10.

[165] Jason W. Buel, "Automated Visions, Algorithmic Imageflows: The Technopolitics of Black Lives Matter Videos on YouTube," *MAST The Journal of Media Art Study and Theory* 3, no. 1 (April 2022): 115.

[166] Nicholas Mirzoeff, "On Visuality," *Journal of Visual Culture* 5 (April 1, 2006): 67.

[167] Mark Fisher, *Capitalist Realism: Is There No Alternative?* (Zero Books, 2009): 9.

[168] Hito Steyerl, "A Sea of Data: Pattern Recognition and Corporate Animism (Forked Version)" in Clemens Apprich et al., *Pattern Discrimination* (Meson Press, University of Minnesota Press, 2018), 5.

[169] Nicholas Mirzoeff, *The Right to Look*: 296, 302.

[170] Ibid., 279.

[172] Ibid., 283.

[173] Ibid., 292.

[174] Ibid.

[176] Ibid., 290.

PART TWO

[1] Wendy Hui Kyong Chun and Alex H. Barnett, *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition* (Cambridge, Massachusetts London, England: The MIT Press, 2021).

[2] Mark Fisher, *Capitalist Realism: Is There No Alternative?* (Zero Books, 2009), 78.

[3] Ganesh Sitaraman, "The Collapse of Neoliberalism," *The New Republic*, December 23, 2019, https://newrepublic.com/article/155970/collapse-neoliberalism.

[4] Fisher, *Capitalist Realism*, 78.

[5] Ibid., 17.

[6] Ibid., 2.

[7] Japhy Wilson and E. Swyngedouw, *The Post-Political and Its Discontents: Spaces of Depoliticisation, Spectres of Radical Politics* (Edinburgh: University Press, 2014).

[8] Bennett and Livingston, "A Brief History of the Disinformation Age," 8.

[9] Ibid.

[11] Ibid., 35.

[12] Edward S. Herman and Noam Chomsky, *Manufacturing Consent: The Political Economy of the Mass Media.*, 1989.

[13] Bennett and Livingston, "A Brief History of the Disinformation Age," 29.

[15] Ibid.

[17] Andreas Jungherr and Ralph Schroeder, "Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy," *Social Media + Society* 7, no. 1 (January 1, 2021): 4.

[18] Joseph Bernstein, "Bad News: Selling the Story of Disinformation," *Harper's Magazine*, accessed March 18, 2024, https://harpers.org/archive/2021/09/bad-news-selling-the-story-of-disinformation/.

[19] Chun, *Discriminating Data*: 82.

[20] Wendy Chun qtd. in Sonja Solomun, "Polarization as the Technological Goal – Not the Error," Centre for Media, Technology and Democracy, accessed August 6, 2024, https://www.mediatechdemocracy.com/all-work/polarization-as-the-technological-goal-not-the-error.

[21] Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You*, 1st ptg (Penguin Press, 2011)

[22] Adam Mosseri, "Bringing People Closer Together," *Meta* (blog), January 12, 2018, https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/.

[23] Jeremy Merrill and Will Oremus, "Five Points for Anger, One for a 'like': How Facebook's Formula Fostered Rage and Misinformation," Washington Post, October 26, 2021, https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/.

[24] Chun qtd. in Solomun, "Polarization as the Technological Goal […]."

[26] Chun, *Discriminating Data*: 85.

[27] Ibid., 243.

[28] Bennett and Livingston, "A Brief History of the Disinformation Age," 20.

[30] Jungherr and Schroeder, "Disinformation and the Structural Transformations of the Public Arena […]," 3.

[32] Fisher, *Capitalist Realism*, 2.

[33] Bennett and Livingston, "A Brief History of the Disinformation Age," 31-32.

[34] Mirzoeff, *The Right to Look*, 303.

[35] Barry Buzan, Ole Wæver, and Jaap de Wilde, *Security: A New Framework for Analysis* (Lynne Rienner Publishers, 1998): 23. [emphasis mine]

[36] Ibid., 26.

[37] Ibid., 24.

[38] Nina Schick, *DEEPFAKES*. "Fucked-up dystopia," Paragraphs 8-11. Ebook.

[39] Adrienne Thompson, "The Existential Threat of AI-Enhanced Disinformation Operations," *Center for Security and Emerging Technology* (blog), July 11, 2022, https://cset.georgetown.edu/article/the-existential-threat-of-ai-enhanced-disinformation-operations/.

[40] "Globe Editorial: Technology and the Law: Catching up to the Existential Threat of Deepfakes," *The Globe and Mail*, June 27, 2024, https://www.theglobeandmail.com/opinion/editorials/article-technology-and-the-law-catching-up-to-the-existential-threat-of/.

[41] *The Clear and Present Danger of Disinformation*, World Economic Forum Annual Meeting (World Economic Forum, 2023), https://www.weforum.org/events/world-economic-forum-annual-meeting-2023/sessions/the-clear-and-present-danger-of-disinformation/.

[42] "'Existential to Who?' US VP Kamala Harris Urges Focus on near-Term AI Risks," POLITICO, November 1, 2023, https://www.politico.eu/article/existential-to-who-us-vp-kamala-harris-urges-focus-on-near-term-ai-risks/.

[43] Buzan et al., *Security* […]: 25.

[45] Francisco Eiras et al., "Near to Mid-Term Risks and Opportunities of Open Source Generative AI" (arXiv, April 25, 2024), http://arxiv.org/abs/2404.17047.

[46] "The Global Risks Report 2024" (World Economic Forum, January 10, 2024), https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf.

[47] "Trends in Internet Use and Attitudes: Findings from a Survey of Canadian Internet Users" (The Strategic Counsel, March 2024), 92-94. https://www.cira.ca/uploads/2024/07/2024-Internet-Trends-Factbook-FINAL-Report.pdf.

[48] Christopher St Aubin and Jacob Liedke, "-Most Americans Favor Restrictions on False Information, Violent Content Online," *Pew Research Center* (blog), July 20, 2023, https://www.pewresearch.org/short-reads/2023/07/20/most-americans-favor-restrictions-on-false-information-violent-content-online/.

[49] Christopher St Aubin and Jacob Liedke, "-Most Americans Favor Restrictions on False Information, Violent Content Online," *Pew Research Center* (blog), July 20, 2023, https://www.pewresearch.org/short-reads/2023/07/20/most-americans-favor-restrictions-on-false-information-violent-content-online/.

[50] Ibid.

[51] Buzan et al., *Security* […]: 31.

[52] Thierry Balzacq, "The Three Faces of Securitization: Political Agency, Audience and Context," European Journal of International Relations 11, no. 2 (June 2005): 179. [emphasis added]

[53] Felix M. Simon, Sacha Altay, and Hugo Mercier, "Misinformation Reloaded? Fears about the Impact of Generative AI on Misinformation Are Overblown," *Harvard Kennedy School Misinformation Review*, October 18, 2023: 4. https://doi.org/10.37016/mr-2020-127.

[54] Alberto Acerbi, Sacha Altay, and Hugo Mercier, "Research Note: Fighting Misinformation or Fighting for Information?," *Harvard Kennedy School Misinformation Review*, January 12, 2022, https://doi.org/10.37016/mr-2020-87.

[55] Yang Cheng and Zifei Fay Chen, "The Influence of Presumed Fake News Influence: Examining Public Support for Corporate Corrective Response, Media Literacy Interventions, and Governmental Regulation," *Mass Communication and Society* 23, no. 5 (September 2, 2020): 705–29, https://doi.org/10.1080/15205436.2020.1750656.

[56] Andreas Jungherr and Adrian Rauchfleisch, "Negative Downstream Effects of Alarmist Disinformation Discourse: Evidence from the United States," *Political Behavior*, January 12, 2024, https://doi.org/10.1007/s11109-024-09911-3.

[57] Jungherr and Schroeder, "Disinformation and the Structural Transformations of the Public Arena […]," 4.

[60] Ibid., 57.

[61] Jack Bratich, "Civil Society Must Be Defended: Misinformation, Moral Panics, and Wars of Restoration," *Communication, Culture and Critique* 13, no. 3 (September 1, 2020): 23.

[64] Bratich, "Civil Society Must Be Defended […]": 312.

[65] Jungherr and Schroeder, "Disinformation and the Structural Transformations of the Public Arena […]," 4.

[66] Bratich, "Civil Society Must Be Defended […]": 314. [emphasis original]

[67] Ibid., 319.

[71] David Kilcullen, *Counterinsurgency* (Oxford University Press, USA, 2010): 2.

[72] Ibid., 215.

[74] Bratich, "Civil Society Must Be Defended […]": 322.

[75] Anton Jäger, "From Post-Politics to Hyper-Politics," Jacobin, February 14, 2022, https://jacobin.com/2022/02/from-post-politics-to-hyper-politics.

PART THREE

[1] Paul Virilio, *The Vision Machine*, Perspectives (Indiana University Press, 1994), 60.

[2] Ibid. [emphasis original]

[4] Trevor Paglen, "Invisible Images (Your Pictures Are Looking at You)," *The New Inquiry* (blog), December 8, 2016, https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/.

[5] Adrian MacKenzie and Anna Munster. "Platform Seeing: Image Ensembles and Their Invisualities." *Theory, Culture & Society 36*, no. 5 (September 1, 2019): 17.

[6] Ibid., 18.

[8] Ibid., 5. [emphasis original]

[9] Paglen, "Invisible Images […]."

[10] Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,* FAccT '21 (New York, NY, USA: Association for Computing Machinery, 2021), 610–23.

[11] Elise Berlinski, Jérémy Morales, and Samuel Sponem, "Artificial Imaginaries: Generative AIs as an Advanced Form of Capitalism," *Critical Perspectives on Accounting* 99 (March 1, 2024): 5.

[12] Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press, 2021), 139. [emphasis mine]

[14] Foucault, *Discipline and Punish* […], 82.

[15] Bender et al., "On the Dangers of Stochastic Parrots […]," 613.

[16] "Common Crawl - Open Repository of Web Crawl Data." Common Crawl - Open Repository of Web Crawl Data. Accessed August 6, 2025. https://commoncrawl.org/.

[17] Jia Deng et al., "ImageNet: A Large-Scale Hierarchical Image Database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, 248–55.

[18] NVIDIA Research Projects. "Flickr-Faces-HQ Dataset (FFHQ)." GitHub Repository. https://github.com/NVlabs/ffhq-dataset/.

[19] David Pereira, "Bias in AI: Much more than a Data problem." *Towards Data Science* (blog), July 6, 2020. https://towardsdatascience.com/bias-in-ai-much-more-than-a-data-problem-de6ef950c848

[20] Tero Karras, Samuli Laine, and Timo Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks" (arXiv, March 29, 2019)

[21] Sachit Menon et al., "PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models" (arXiv, July 20, 2020),

[22] Vongani H. Maluleke et al., "Studying Bias in GANs through the Lens of Race" (arXiv, September 14, 2022), 7.

[23] Alexandra Sasha Luccioni et al., "Stable Bias: Analyzing Societal Representations in Diffusion Models" (arXiv, November 9, 2023)

[24] Shangbin Feng et al., "From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ed. Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (ACL 2023, Toronto, Canada: Association for Computational Linguistics, 2023), 11737.

[25] Fabian Offert and Thao Phan, "A Sign That Spells: DALL-E 2, Invisual Images and The Racial Politics of Feature Space" (arXiv, October 26, 2022), 3. [emphasis mine]

[26] Sylvia Wynter, "UNPARALLELED CATASTROPHE FOR OUR SPECIES? Or, to Give Humanness a Different Future: Conversations," in *Sylvia Wynter: On Being Human as Praxis* (Duke University Press, 2015), 30. [emphasis original]

[27] Crawford, *Atlas of AI*, 184.

[28] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," *Nature* 521, no. 7553 (May 2015): 436. [emphasis mine]

[29] Marco Donnarumma, "Against the Norm: Othering and Otherness in AI Aesthetics," *Digital Culture & Society 8,* no. 2 (December 1, 2022): 11.

[31] Hongbin Ye et al., "Cognitive Mirage: A Review of Hallucinations in Large Language Models" (arXiv, September 13, 2023)

[32] Louise Amoore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others* (Durham: Duke University Press, 2020), 69. [emphasis added]

[34] Ibid., 73.

[35] David Kilcullen, *Counterinsurgency* (Oxford University Press, USA, 2010), 2.

[36] Nicholas Mirzoeff, *The Right to Look: A Counterhistory of Visuality* (Durham: Duke University Press, 2011), 280.

[37] Adam Harvey, "Circular Diffusion," *Adam Harvey Studio* (blog), August 21, 2022, https://adam.harvey.studio/circular-diffusion/.

[38] Gary Marcus, "What If Generative AI Turned out to Be a Dud?," Substack newsletter, *Marcus on AI* (blog), August 12, 2023, https://garymarcus.substack.com/p/what-if-generative-ai-turned-out.

[39] NVIDIA. "NVIDIA Keynote at SIGGRAPH 2023," Youtube. Aug. 8 2023. 20:00-20:42. https://www.youtube.com/watc,h?v=Z2VBKerS63A. [emphasis mine]

[40] Ibid.

[41] Roman Krzanowski and Pawel Polak, "The Internet as an Epistemic Agent (EA)," *Információs Társadalom* 22, no. 2 (August 31, 2022): 44.

[42] Ibid., 45.

[43] Meta. "What's New Across Our AI Experiences," Meta (blog), December 6, 2023, https://about.fb.com/news/2023/12/meta-ai-updates/.

[44] Jared Spataro, "Introducing Microsoft 365 Copilot – Your Copilot for Work," *The Official Microsoft Blog*, March 16, 2023, https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/. [emphasis original]

[45] Apple. "Apple Unveils M3, M3 Pro, and M3 Max, the Most Advanced Chips for a Personal Computer," *Apple Newsroom* (Canada), accessed February 13, 2024, https://www.apple.com/ca/newsroom/2023/10/apple-unveils-m3-m3-pro-and-m3-max-the-most-advanced-chips-for-a-personal-computer/.

[46] Anshel Sag, "AI Dominates Qualcomm Snapdragon Summit With New Snapdragon Products," *Forbes*, accessed February 13, 2024, https://www.forbes.com/sites/moorinsights/2023/10/25/ai-dominates-qualcomm-snapdragon-summit-with-new-snapdragon-products/.

[47] Srinivasan Venkatachary et al., "A New Way to Search with Generative AI: An Overview of SGE" (Alphabet, January 2024): 3.

[48] Shahan Ali Memon and Jevin D. West, "Search Engines Post-ChatGPT: How Generative Artificial Intelligence Could Make Search Less Reliable" (arXiv, February 18, 2024), 4.

[49] Felix M. Simon, "Artificial Intelligence in the News: How AI Retools, Rationalizes, and Reshapes Journalism and the Public Arena" (Tow Center for Digital Journalism: Columbia University, 2024), 35.

[50] Miles Kruppa, "Jeff Bezos Bets on a Google Challenger Using AI to Try to Upend Internet Search," *WSJ*, January 4, 2024, https://www.wsj.com/tech/ai/jeff-bezos-bets-on-a-google-challenger-using-ai-to-try-to-upend-internet-search-0859bda6.

[51] Shahan Ali Memon and Jevin D. West, "Search Engines Post-ChatGPT: How Generative Artificial Intelligence Could Make Search Less Reliable" (arXiv, February 18, 2024), 6. [emphasis mine]

[53] Helen Nissenbaum and Lucas D. Introna, "Shaping the Web: Why the Politics of Search Engines Matters," *The Information Society* 16, no. 3 (July 1, 2000): 169–85.

[54] Rand Fishkin, "An Anonymous Source Shared Thousands of Leaked Google Search API Documents with Me; Everyone in SEO Should See Them," SparkToro, May 28, 2024, https://sparktoro.com/blog/an-anonymous-source-shared-thousands-of-leaked-google-search-api-documents-with-me-everyone-in-seo-should-see-them/.

[55] "Google's AI Search Setback," Platformer, May 29, 2024, https://www.platformer.news/google-ai-overviews-eat-rocks-glue-pizza/.

[56] Benjamin N Jacobsen, "Regimes of Recognition on Algorithmic Media," *New Media & Society* 25, no. 12 (December 1, 2023): 3652.

[57] Rishi Bommasani et al., "On the Opportunities and Risks of Foundation Models" (arXiv, July 12, 2022), 7.

[58] Ibid., 4.

[59] "GPT-4." OpenAI. Accessed February 24, 2024. https://openai.com/index/gpt-4/.

[60] Alex Williams, "Control Societies and Platform Logic.('Postscript on the Societies of Control' by Gilles Deleuze)," *New Formations*, no. 84–85 (2015): 220.

[61] Ibid., 222.

[62] David M. Berry and James Stockman, "Schumacher in the Age of Generative AI: Towards a New Critique of Technology," *European Journal of Social Theory*, March 1, 2024: 11.

[63] Bommasani et al., "On the Opportunities and Risks of Foundation Models": 5. [emphasis original and added]

[65] Alexander R. Galloway, *Protocol: How Control Exists after Decentralization*, Leonardo (Cambridge, Mass: MIT Press, 2004), 7.

[66] Andreas Jungherr and Ralph Schroeder, "Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy," *Social Media + Society* 7, no. 1 (January 1, 2021): 5.

[67] Williams, "Control Societies and Platform Logic […]" : 225. [emphasis mine]

[68] Keith Raymond Harris, "Epistemic Domination," *Thought: A Journal of Philosophy*, 2023, https://doi.org/10.5840/tht202341317.

[69] Ibid., 4.

[70] Ibid., 1.

[71] Zachary P Devereaux et al., "Contemporary Media and the AI-Goldrush from a Military Perspective." (2023 IDeaS Conference: Disinformation, Hate Speech, and Extremism Online, Carnegie Mellon Univeristy, 2023).

[72] Galloway, *Protocol: How Control Exists after Decentralization*: 75.

[73] Fred Bauer, "Why We Must Resist AI's Soft Mind Control," *The Atlantic*, March 2024, https://www.theatlantic.com/ideas/archive/2024/03/artificial-intelligence-google-gemini-mind-control/677683/

[74] Jacques Ellul. *Propaganda* […]: 11, 31.

[76] Ibid., 25.

[78] Kilcullen, *Counterinsurgency*: 208. [emphasis mine]

[79] Ibid., x.

[80] Ibid., 216.

[81] Williams, "Control Societies and Platform Logic […]" : 214.

[83] Kilcullen, *Counterinsurgency*: 42.

[84] Ibid., 30. [emphasis mine]

[85] Kilcullen, *Counterinsurgency*: 194. [emphasis original and added]

[87] Ibid., 167.

[88] Mirzoeff, *The Right to Look*, 3.

[89] Ibid., 208.

[90] Ibid., 215.

[91] Ibid., 42.

[92] Harris, "Epistemic Domination," 5.

[93] "The Hiroshima AI Process: Leading the Global Challenge to Shape Inclusive Governance for Generative AI," The Government of Japan, February 9, 2024, https://www.japan.go.jp/kizuna/2024/02/hiroshima_ai_process.html.

[94] "Bletchley Park Makes History Again as Host of the World's First AI Safety Summit," Bletchley Park, November 1, 2023, https://bletchleypark.org.uk/bletchley-park-makes-history-again-as-host-of-the-worlds-first-ai-safety-summit/.

[95] "High-Level Summary of the AI Act," *EU Artificial Intelligence Act* (blog), February 27, 2024, https://artificialintelligenceact.eu/high-level-summary/.

[97] Council of the European Union, *EU Artificial Intelligence Act*, 26 January 2024, 63, https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf

[99] Ibid., 5.

[100] Ibid., 81.

[101] Supantha Mukjerjee, "Explainer: Will the EU Delay Enforcing Its AI Act?," Reuters, July 3, 2025, https://www.reuters.com/business/media-telecom/will-eu-delay-enforcing-its-ai-act-2025-07-03/.

[102] "General-Purpose AI Code of Practice Now Available," European Commission, July 9, 2025, https://ec.europa.eu/commission/presscorner/detail/en/ip_25_1787.

[103] Central Committee of the Communist Party of China, "New Information Network Technology" in *The 13th Five-Year Plan for Economic and Social Development of the People's Republic of China 2016-2020*, https://en.ndrc.gov.cn/policies/202105/P020210527785800103339.pdf ; and Central Committee of the Communist Party of China, "Encouraging ground-breaking scientific and technological innovation" in *Outline of the 14th Five-Year Plan (2021-2025) for National Economic and Social Development and Vision 2035 of the People's Republic of China*, https://www.fujian.gov.cn/english/news/202108/t20210809_5665713.htm

[104] State Council of China, *Notice of the Development Plan for the New Generation of Artificial Intelligence*, July 2017, https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

[105] State Internet Information Office, *Internet Information Service Algorithm Recommendation Management Regulations*, December 2021, https://www.gov.cn/zhengce/zhengceku/2022-01/04/content_5666429.htm

[106] State Internet Information Office, *Internet Service Deep Synthesis Management Provisions*, November 2022, https://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm

[107] State Internet Information Office, *Interim Measures for Generative Artificial Intelligence Service Management*, July 2023, https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

[108] Ibid., Article 1.

[109] Ibid. Article 4(1).

[111] Ibid., Article 14.

[112] State Internet Information Office, *Interim Measures for Generative Artificial Intelligence Service Management*, April 2023, (draft), Article 9. https://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm

[113] State Internet Information Office, *Interim Measures for Generative Artificial Intelligence* […], Article 20.

[114] "Executive Order 14110 of October 20, 2023, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," *Code of Federal Regulations*, 2023-24283 (2023): 75191-75226, https://www.federalregister.gov/d/2023-24283

[115] Ibid., Section 2(a).

[116] Ibid., Section 3(d).

[118] Ibid., Section 10.1(b)(viii)(A).

[119] Ibid., Section 4.2(b).

[120] "U.S. Artificial Intelligence Safety Institute." National Institute of Standards and Technology, February 8, 2024. https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute.

[121] Robert Chesney and Danielle Keats Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," *SSRN Electronic Journal*, 2018: 1804-1807.

[122] Nicol Turner Lee and Obioha Chijioke, "Why States and Localities Are Acting on AI," Brookings, accessed March 15, 2024, https://www.brookings.edu/articles/why-states-and-localities-are-acting-on-ai/.

[123] "America's AI Action Plan" (Executive Office of the President of the United States, July 2025), https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.

[124] Ibid., 4.

[125] Barry Friedman, "What Is Public Safety?," *Boston University Law Review* 102, no. 3 (2022): 725–92.

[126] David Baldridge, Beth Coleman, and Jamie Sandhu, "The Terminology of AI Regulation: Preventing 'Harm' and Mitigating 'Risk,'" Schwartz Reisman Institute, accessed March 18, 2024, https://srinstitute.utoronto.ca/news/terminology-regulation-risk-harm.

[127] Council of the European Union, *EU Artificial Intelligence Act*: Article 3(44)(d).

[128] Joshua A. Tucker et al., "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature," SSRN Scholarly Paper (Rochester, NY, March 19, 2018): 55. https://doi.org/10.2139/ssrn.3144139.

[129] Ladislav Bittman, *The KGB and Soviet Disinformation: An Insider's View* (Pergamon-Brassey's, 1985), 44.

[130] Joseph Bernstein, "Bad News: Selling the Story of Disinformation," *Harper's Magazine*, accessed March 18, 2024, https://harpers.org/archive/2021/09/bad-news-selling-the-story-of-disinformation/.

[132] Kilcullen, *Counterinsurgency*: 215.

[133] Ingrid Lunden, "UK Drops 'safety' from Its AI Body, Now Called AI Security Institute, Inks MOU with Anthropic," *TechCrunch* (blog), February 14, 2025, https://techcrunch.com/2025/02/13/uk-drops-safety-from-its-ai-body-now-called-ai-security-institute-inks-mou-with-anthropic/.

[134] "Statement from U.S. Secretary of Commerce Howard Lutnick on Transforming the U.S. AI Safety Institute into the Pro-Innovation, Pro-Science U.S. Center for AI Standards and Innovation," US Department of Commerce, June 3, 2025, https://www.commerce.gov/news/press-releases/2025/06/statement-us-secretary-commerce-howard-lutnick-transforming-us-ai.

[135] A. Burns, J. McDermid, and J. Dobson, "On the Meaning of Safety and Security," *The Computer Journal* 35, no. 1 (February 1, 1992): 3.

[136] Ibid., 4.

[137] National Security Council, "National Strategy for Countering Domestic Terrorism," June 2021, 8. https://www.whitehouse.gov/wp-content/uploads/2021/06/National-Strategy-for-Countering-Domestic-Terrorism.pdf

[138] "Statement from U.S. Secretary of Commerce Howard Lutnick […],".

[139] Faiza Patel, "U.S. AI-Driven 'Catch and Revoke' Initiative Threatens First Amendment Rights | Brennan Center for Justice," Brennan Center for Justice, March 18, 2025, https://www.brennancenter.org/our-work/analysis-opinion/us-ai-driven-catch-and-revoke-initiative-threatens-first-amendment-rights.

[140] Ken Dilanian and Julia Ainsley, "DHS Weighing Major Changes to Fight Domestic Violent Extremism," NBC News, March 25, 2021, https://www.nbcnews.com/politics/national-security/dhs-weighing-huge-changes-fight-domestic-violent-extremism-say-officials-n1262047.

[141] Ibid., 7.

[143] Gillian Hadfield, Mariano-Florentino Cuéllar, Tim O'Reilly, "It's Time to Create a National Registry for Large AI Models," Carnegie Endowment for International Peace, accessed March 21, 2024, https://carnegieendowment.org/2023/07/12/it-s-time-to-create-national-registry-for-large-ai-models-pub-90180.

[145] Blake Brittain, "How Copyright Law Could Threaten the AI Industry in 2024," Reuters, accessed March 21, 2024, https://www.reuters.com/legal/litigation/how-copyright-law-could-threaten-ai-industry-2024-2024-01-02/.

[146] Department of Commerce, "Implementation of Additional Export Controls: Certain Advanced Computing Items; Supercomputer and Semiconductor End Use; Updates and Corrections," *Federal Registry* 0694-AI94, October 25, 2023. https://www.federalregister.gov/d/2023-23055

[147] Jiaming Ji et al., "AI Alignment: A Comprehensive Survey" (arXiv, February 26, 2024), 3. http://arxiv.org/abs/2310.19852.

[148] Ibid., 9.

[149] Ibid.

[151] Ibid., 13.

[152] "Common Crawl - Open Repository of Web Crawl Data," accessed April 2, 2024, https://commoncrawl.org/.

[153] Colin Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer" (arXiv, September 19, 2023), 6. http://arxiv.org/abs/1910.10683.

[154] "List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words," GitHub, accessed April 2, 2024, https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words/blob/master/en.

[155] Chinese National Information Security Standardization Technical Committee, Ben Murphy trans. *Basic Safety Requirements for Generative Artificial Intelligence Services (Draft for Feedback)*, Oct. 2023, https://cset.georgetown.edu/publication/china-safety-requirements-for-generative-ai/

[156] State Internet Information Office, *Provisions on the Governance of the Online Information Content Ecosystem*, December 2019, Articles 6 & 7. https://www.cac.gov.cn/2019-12/20/c_1578375159509309.htm

[157] Dieuwke Hupkes et al., "A Taxonomy and Review of Generalization Research in NLP," *Nature Machine Intelligence* 5, no. 10 (October 2023): 1161–74.

[159] "Aligning Language Models to Follow Instructions," OpenAI, January 27, 2022, https://openai.com/research/instruction-following.

[160] Long Ouyang et al., "Training Language Models to Follow Instructions with Human Feedback" (arXiv, March 4, 2022), 2. http://arxiv.org/abs/2203.02155.

[161] "Introducing Superalignment," OpenAI, July 5, 2023, https://openai.com/blog/introducing-superalignment

[162] "Model Card and Evaluations for Claude Models," Anthropic, July 8, 2023. https://www-cdn.anthropic.com/bd2a28d2535bfb0494cc8e2a3bf135d2e7523226/Model-Card-Claude-2.pdf

[163] Yuntao Bai et al., "Constitutional AI: Harmlessness from AI Feedback" (arXiv, December 15, 2022), 1. https://doi.org/10.48550/arXiv.2212.08073.

[166] Bai et al., "Constitutional AI: Harmlessness from AI Feedback," 22.

[167] "Claude's Constitution," Anthropic, May 9, 2023. https://www.anthropic.com/news/claudes-constitution

[168] Zhiyuan Yu et al., "Don't Listen To Me: Understanding and Exploring Jailbreak Prompts of Large Language Models" (arXiv, March 25, 2024), https://doi.org/10.48550/arXiv.2403.17336.

[169] "Moderations" OpenAI Platform, accessed April 8, 2024, https://platform.openai.com.docs/api-reference/moderation

[170] "DallE-3 System Card," OpenAI, October 3, 2023, https://cdn.openai.com/papers/DALL_E_3_System_Card.pdf

[171] Todor Markov et al., "A Holistic Approach to Undesired Content Detection in the Real World" (arXiv, February 14, 2023), http://arxiv.org/abs/2208.03274.

[172] Johannes Welbl et al., "Challenges in Detoxifying Language Models," in *Findings of the Association for Computational Linguistics: EMNLP 2021*, ed. Marie-Francine Moens et al. (Findings 2021, Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021), 2447–69, https://doi.org/10.18653/v1/2021.findings-emnlp.210.

[173] Patrick Schramowski et al., "Safe Latent Diffusion: Mitigating Inappropriate Degeneration in Diffusion Models" (arXiv, April 26, 2023), 6. http://arxiv.org/abs/2211.05105.

[174] Ibid., 6.

[175] "AIML-TUDA/I2p," Datasets at Hugging Face, accessed April 8, 2024, https://huggingface.co/datasets/AIML-TUDA/i2p.

[176] Patrick Schramowski et al., "Safe Latent Diffusion," 6.

[177] "A List of Leaked System Prompts," Blog, Matt Rickard, May 17, 2023, https://matt-rickard.com/a-list-of-leaked-system-prompts.

[178] Matthias Bastian, "DALL-E 3's System Prompt Reveals OpenAI's Rules for AI Image Generation," THE DECODER, October 16, 2023, https://the-decoder.com/dall-e-3s-system-prompt-reveals-openais-rules-for-generative-image-ai/.

[179] "DallE-3 System Card," OpenAI, 2.

[180] James Betker et al., "Improving Image Generation with Better Captions," OpenAI, October 2023, 9.

[182] Jake Traylor, "No Quick Fix: How OpenAI's DALL·E 2 Illustrated the Challenges of Bias in AI," NBC News, July 27, 2022, https://www.nbcnews.com/tech/tech-news/no-quick-fix-openais-dalle-2-illustrated-challenges-bias-ai-rcna39918.

[183] Benj Edwards, "Google's Hidden AI Diversity Prompts Lead to Outcry over Historically Inaccurate Images," Ars Technica, February 22, 2024, https://arstechnica.com/information-technology/2024/02/googles-hidden-ai-diversity-prompts-lead-to-outcry-over-historically-inaccurate-images/.

[184] Matt Growcoot, "Meta's AI Image Generator Accused of Making 'Woke' Pictures Like Google," PetaPixel, March 4, 2024, https://petapixel.com/2024/03/04/metas-ai-image-generator-accused-of-making-woke-pictures-like-google/.

[185] Alexander R. Galloway, *Protocol: How Control Exists after Decentralization*, Leonardo (Cambridge, Mass: MIT Press, 2004), 75.

[186] Ibid., 53.

[187] Britt Paris, "Seeing Through the Fog of War: Assessing Epistemic Burden Around Cheapfakes and Deepfakes of Geopolitical Crisis," in *Re-Thinking Mediations of Post-Truth Politics and Trust* (Routledge, 2023), 154.

[188] Williams, "Control Societies and Platform Logic […]," 220.

[190] Wendy Hui Kyong Chun and Alex H. Barnett, *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition* (Cambridge, Massachusetts London, England: The MIT Press, 2021), 41.

[191] Maximilian Kasy, "The Political Economy of AI: Towards Democratic Control of the Means of Prediction," Discussion Paper Series (IZA Institute of Labour Economics, April 2024): 21.

[192] Emma J Llansó, "No Amount of 'AI' in Content Moderation Will Solve Filtering's Prior-Restraint Problem," *Big Data & Society* 7, no. 1 (January 1, 2020): 3. [emphasis original]

[193] Jordi Calvet-Bademunt and Jacob Mchangama, "Freedom of Expression in Generative AI: A Snapshot of Content Policies" (The Future of Free Speech, February 2024), 20.

[195] Ibid.

[196] Ibid., 30.

[197] Kilcullen, *Counterinsurgency*: 7.

[198] "DARPA Announces Research Teams Selected to Semantic Forensics Program," March 2, 2021, https://www.darpa.mil/news-events/2021-03-02.

[199] Deep Media, "Deep Media AI," Accessed April 20, 2024. https://deepmedia.ai/

[200] Frank Wolfe, "San Francisco-Based Firm Announces DoD SBIR Contracts for Deep Fake Detection," Defense Daily, November 28, 2022, https://www.defensedaily.com/san-francisco-based-firm-announces-dod-sbir-contracts-for-deep-fake-detection/air-force/.

[201] "The Best Deepfake Detection Solution | AI Image Detection," Sensity AI, Accessed April 20, 2024. https://sensity.ai/deepfakes-detection/

[202] "Deepware | Scan and Detect Deepfake Videos," Deepware. https://deepware.ai/

[203] "MeVer," MeVer. Accessed April 20, 2024. https://mever.gr/services/

[204] "Technology — Reality Defender," Reality Defender. Accessed April 20, 2024. https://realitydefender.com/technology

[205] Jan Kirchner et al., "New AI Classifier for Indicating AI-Written Text," accessed April 24, 2024, https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text.

[206] Benj Edwards, "OpenAI Confirms That AI Writing Detectors Don't Work," Ars Technica, September 8, 2023, https://arstechnica.com/information-technology/2023/09/openai-admits-that-ai-writing-detectors-dont-work/.

[207] The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI," The White House, July 21, 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

[208] "C2PA Founding Press Release - C2PA," Coalition for Content Provenance and Authenticity, February 22, 2021, https://c2pa.org/post/c2pa_initial_pr/.

[209] "Introducing the Content Authenticity Initiative," Content Authenticity Initiative, November 4, 2019, https://contentauthenticity.org/blog/test.

[210] J. Aythora et al., "MULTI-STAKEHOLDER MEDIA PROVENANCE MANAGEMENT TO COUNTER SYNTHETIC MEDIA RISKS IN NEWS PUBLISHING" (International Broadcasting Conference, virtual, 2020), https://drive.google.com/file/d/11c41iTwq7z-nMMMPQLE5GZ6gJmc_llJ1/view?usp=sharing&usp=embed_facebook.

[211] "About — C2PA," Coalition for Content Provenance and Authenticity, https://c2pa.org/about/about/

[212] "Content Credentials : C2PA Technical Specification :: C2PA Specifications," accessed April 2, 2024, https://c2pa.org/specifications/specifications/2.0/specs/C2PA_Specification.html#_technical_overview.

[213] Ibid., "Multiple Claims."

[214] Ibid., "Establishing Trust."

[215] Neal Krawetz, "C2PA's Butterfly Effect,' Hacker Factor, November 16, 2023, accessed February 26, 2024, https://www.hackerfactor.com/blog/index.php?/archives/1010-C2PAs-Butterfly-Effect.html

[216] Neal Krawetz, "C2PA's Worst Case Scenario,' Hacker Factor, December 18, 2023, accessed February 26, 2024, https://www.hackerfactor.com/blog/index.php?/archives/1013-C2PAs-Worst-Case-Scenario.html

[217] "13.1. Overview," C2PA Technical Specification, accessed April 2, 2024, https://c2pa.org/specifications/specifications/1.0/specs/C2PA_Specification.html#_trust_model [emphasis added]

[218] "3.2. Trust," C2PA Explainer, accessed May 4, 2024, https://c2pa.org/specifications/specifications/1.3/explainer/Explainer.html#_trust

[219] "6.3. General considerations for civic, community, and independent media," C2PA Harms Modelling, accessed May 6, 2024, https://c2pa.org/specifications/specifications/1.0/security/Harms_Modelling.html#_general_considerations_for_content_creators [emphasis added]

[221] J. Aythora et al., "MULTI-STAKEHOLDER MEDIA PROVENANCE MANAGEMENT TO COUNTER SYNTHETIC MEDIA RISKS IN NEWS PUBLISHING," 3. [emphasis added]

[222] "5.3.2.4. Social Media and Video Sharing," C2PA Implementation Guidance, accessed May 1, 2024, https://c2pa.org/specifications/specifications/1.2/guidance/Guidance.html#sharing-sites

[223] K. J. Kevin Feng et al., "Examining the Impact of Provenance-Enabled Media on Trust and Accuracy Perceptions," *Proceedings of the ACM on Human-Computer Interaction* 7, no. CSCW2 (October 4, 2023): 3.

[224] Megan Crouse, "Could C2PA Cryptography Be the Key to Fighting AI-Driven Misinformation?," TechRepublic, August 3, 2023, https://www.techrepublic.com/article/ai-cryptography-watermarking/.

[225] TruePic. "C2PA Digital Content Provenance Virtual Event," Youtube, 2023, https://www.youtube.com/watch?v=vrrA_UJTZHM.

[227] Gordon Pennycook et al., "The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings," *Management Science* 66, no. 11 (November 2020): 4944–57, https://doi.org/10.1287/mnsc.2019.3478.

[228] Harris, "Epistemic Domination," 5.

[230] Matthias Leese, "'Seeing Futures': Politics of Visuality and Affect," in *Algorithmic Life* (Routledge, 2016). 149

[231] Ibid., 152.

[234] Ibid.

[235] "15.2. Identity of Signers," C2PA Specifications. [emphasis added]

[237] "Using C2PA enabled tools and software," C2PA Harms Modelling.

[238] Ben Gomes, "Our Latest Quality Improvements for Search," Google, April 25, 2017, https://blog.google/products/search/our-latest-quality-improvements-search/.

[239] Kilcullen, *Counterinsurgency*, 42.

[240] Britt Paris and Joan Donovan, "Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence," *Data & Society*, September 2019: 22.

[241] Ibid.

[243] Ibid., 7.

[244] Foucault, *Power/Knowledge*: 131.

[245] Regina Rini, "Deepfakes and the Epistemic Backstop," *Philosopher's Imprint* 20, no. 24 (2020)

[246] Yochai Benkler, Robert Faris, and Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (Oxford University Press, 2018), 24.

[247] Tagg, *The Burden of Representation*, 99.

[248] Mirzoeff, *The Right to Look*, 279.

[249] Ibid.

[250] "Christchurch Call" Christchurch Call, accessed May 28, 2024, https://www.christchurchcall.com

[251] "Christchurch Call Leaders' Summit 2023: Compilation of supporting papers," (Christchurch Call Foundation, 2023), https://www.christchurchcall.com/assets/Documents/Christchurch-Call-Leaders-Summit-2023-Supporting-papers.pdf

[252] "Christchurch Call Text," Christchurch Call, accessed May 28, 2024, https://www.christchurchcall.com/about/christchurch-call-text. [emphasis added]

[253] Ministère de l'Europe et des Affaires étrangères, "Joint Communiqué Issued by the President of the French Republic and the Prime Minister of New Zealand: Significant Global Progress Made under Christchurch Call (13 April 2021)," Ministry for Europe and Foreign Affairs, accessed May 28, 2024, https://www.diplomatie.gouv.fr/en/french-foreign-policy/digital-diplomacy/news/article/joint-communique-issued-by-the-president-of-the-french-republic-and-the-prime.

[254] "New Technologies and the Christchurch Call: Challenges, Mitigations, and Opportunities" (Christchurch Call Foundation, 2023), https://www.christchurchcall.com/assets/Summit-23/Leaders-Summit-2023-Meeting-paper-New-technologies-and-the-Christchurch-Call-Challenges-mitigations-and-opportunities.pdf.

[255] "Four New Tech Firms Expand the Christchurch Call," Christchurch Call, accessed May 28, 2024, https://www.christchurchcall.com/media-and-resources/news-and-updates/four-new-tech-firms-join-the-christchurch-call.

[256] "Understanding Algorithms and Developing Interventions," Christchurch Call, accessed May 29, 2024, https://www.christchurchcall.com/our-work/algorithms-and-positive-interventions.

[258] Ibid.

[259] Saadia Gabriel et al., "Generative AI in the Era of 'Alternative Facts,'" *An MIT Exploration of Generative AI*, March 27, 2024, https://doi.org/10.21428/e4baedd9.82175d26.

[260] Francesco Salvi et al., "On the Conversational Persuasiveness of Large Language Models: A Randomized Controlled Trial" (arXiv, March 21, 2024), http://arxiv.org/abs/2403.14380.

[261] Thomas H. Costello, Gordon Pennycook, and David Rand, "Durably Reducing Conspiracy Beliefs through Dialogues with AI" (OSF, April 3, 2024), 13. https://doi.org/10.31234/osf.io/xcwdn.

[262] Maurice Jakesch et al., "Co-Writing with Opinionated Language Models Affects Users' Views," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23 (New York, NY, USA: Association for Computing Machinery, 2023), 1–15, https://doi.org/10.1145/3544548.3581196.

[263] "Defense Department Adoption of Generative Artificial Intelligence," Special Competitive Studies Project, accessed May 29, 2024, https://www.scsp.ai/reports/gen-ai/defense/.

[264] Ronald Watkins, "USSOCOM to Use AI to Detect Disinformation Threats on Social Media," The Defense Post, August 31, 2023, https://www.thedefensepost.com/2023/08/31/ussocom-ai-disinformation-social-media/.

[265] "Meet Argus: Accrete's Advanced AI Threat Detection Agent," Accrete AI, accessed May 29, 2024, https://www.accrete.ai/argus.

[266] Watkins, "USSOCOM to Use AI […]"

[267] "Nebula Social™ by Accrete.AI," Nebula Social, accessed May 29, 2024, https://www.nebulasocial.ai/.

[268] "Insider Q&A: CIA's Chief Technologist's Cautious Embrace of Generative AI," AP News, May 20, 2024, https://apnews.com/article/cia-chief-technologist-gen-ai-c3f4154a8e2879d7820931f07656968c.

[269] Mirzoeff, *The Right to Look*, 295.

[270] Nicholas Mirzoeff, "On Visuality," *Journal of Visual Culture* 5 (April 1, 2006): 57.

[272] Alexander R. Galloway, *Gaming: Essays on Algorithmic Culture*, Electronic Mediations ; v. 18 (Minneapolis: University of Minnesota Press, 2006), 63.

[273] Mirzoeff, *The Right to Look*, 296.

[274] Ibid., 297.

[275] "Meet Argus: Accrete's Advanced AI Threat Detection Agent," ," Accrete AI, accessed June 25, 2025, https://www.accrete.ai/argus. [emphasis added].

[276] "Anthropic Awarded $200M DOD Agreement for AI Capabilities," Anthropic, July 14, 2025, https://www.anthropic.com/news/anthropic-and-the-department-of-defense-to-advance-responsible-ai-in-defense-operations.; Lucy Hooker and Chris Vallance, "Google Ending AI Arms Ban Incredibly Concerning, Campaigners Say," BBC News, February 5, 2025, https://www.bbc.com/news/articles/cy081nqx2zjo.; Sam Biddle, "OpenAI Quietly Deletes Ban on Using ChatGPT for 'Military and Warfare,'" *The Intercept* (blog), January 12, 2024, https://theintercept.com/2024/01/12/open-ai-military-ban-chatgpt/.

[277] Matthew Olay, "Senior Officials Outline President's Proposed FY26 Defense Budget," *DoD News* (blog), June 26, 2025, https://www.defense.gov/News/News-Stories/Article/Article/4227847/senior-officials-outline-presidents-proposed-fy26-defense-budget/.

[278] Mike Stone and Aishwarya Jain, "Defense Firm Anduril Partners with OpenAI to Use AI in National Security Missions," Reuters, December 4, 2024, https://www.reuters.com/technology/artificial-intelligence/defense-firm-anduril-partners-with-openai-use-ai-national-security-missions-2024-12-04/.

[279] Kurt Wagner, "Meta Opens Its AI Models to US Defense Agencies and Contractors," *Bloomberg.Com*, November 4, 2024, https://www.bloomberg.com/news/articles/2024-11-04/meta-opens-llama-ai-models-to-us-defense-agencies-contractors.

[280] "US Defense Department Awards Contracts to Google, Musk's xAI," *Reuters*, July 14, 2025, sec. Autos & Transportation, https://www.reuters.com/business/autos-transportation/us-department-defense-awards-contracts-google-xai-2025-07-14/.

[281] "Contracts for June 16, 2025," U.S. Department of Defense, June 16, 2025, https://www.defense.gov/News/Contracts/Contract/Article/4218062//.

[282] "OpenAI Appoints Retired U.S. Army General," *OpenAI | Security* (blog), June 13, 2024, https://openai.com/index/openai-appoints-retired-us-army-general/.

[283] US Army Public Affairs, "Army Launches Detachment 201: Executive Innovation Corps to Drive Tech Transformation," www.army.mil, June 13, 2025,

https://www.army.mil/article/286317/army_launches_detachment_201_executive_innovation_corps_to_drive_tech_transformation.

[284] John Feffer, "The De-Trumpification of America - FPIF," *Foreign Policy In Focus* (blog), June 26, 2020, https://fpif.org/the-de-trumpification-of-america/.

[285] In a 2021 appearance on the *Jack Murphy Live* podcast, current US vice president J.D. Vance called for the "de-institutionalization" of left-leaning culture and politics: "We need like a de-Baathification, a de-wokeification program in the United States." – "J.D. Vance - JML #070," posted September 17, 2021 by Jack Murphy Live, Youtube, 01:32:00, https://www.youtube.com/watch?v=PMq1ZEcyztY.

[286] Ibid., 279.

[287] David M. Berry and James Stockman, "Schumacher in the Age of Generative AI: Towards a New Critique of Technology," *European Journal of Social Theory*, March 1, 2024, 11.

[288] Foucault, *Discipline and Punish*, 82.

[289] Ibid., 106.

[290] Berry and Stockman, "Schumacher in the Age of Generative AI […]," 10.

[291] Hilla Dayan, "Regimes of separation: Israel/Palestine and the shadow of apartheid," in *The Power of Inclusive Exclusion: Anatomy of Israeli Rule in the Occupied Palestinian Territories*. Eds. Adi Ophir, M. Givoni, and Sārī Ḥanafī (New York: Zone Books, 2009).

[292] Hilla Dayan qtd. In Mirzoeff, *The Right to Look*, 301.

[293] Tagg, *The Burden of Representation*, 20.

[294] Ibid., 17.

[295] Ibid.

[296] Ibid., 18.

[297] Ibid.

[298] Ibid., 19.

[299] Ibid., 20.

[301] Mirzoeff, *The Right to Look*, 294. [emphasis added]

[302] Ibid., 1.

[303] Ibid., 4.

[304] Jeremy Howard, "AI Safety and the Age of Dislightenment," fast.ai, July 10, 2023, https://www.fast.ai/posts/2023-11-07-dislightenment.html

[305] Mirzoeff, *The Right to Look*, 1. [emphasis added]

[306] Ibid., 2.

CODA

[1] Adrian Shahbaz, "The Rise of Digital Authoritarianism," Freedom House, accessed October 25, 2024, https://freedomhouse.org/report/freedom-net/2018/rise-digital-authoritarianism.

[2] Simeon Gilding, "De-Risking Authoritarian AI: A Balanced Approach to Protecting Our Digital Ecosystems," Policy Brief (Australian Strategic Policy Institute, 2023).

[3] Jennifer Bouey et al., "China's AI Exports: Technology Distribution and Data Safety" (RAND Corporation, December 11, 2023), https://www.rand.org/pubs/research_reports/RRA2696-2.html.

[4] Sam Bresnick, "China Bets Big on Military AI," CEPA, April 3, 2024, https://cepa.org/article/china-bets-big-on-military-ai/.

[5] "China Is Writing the World's Technology Rules," *The Economist*, October 10, 2024, https://www.economist.com/business/2024/10/10/china-is-writing-the-worlds-technology-rules.

[6] State Internet Information Office, *Global AI Governance Initiative*, Office of the Central Cyberspace Affairs Commission, October 2023, https://www.cac.gov.cn/2023-10/18/c_1699291032884978.htm

[7] Department of Commerce, "Implementation of Additional Export Controls: Certain Advanced Computing Items; Supercomputer and Semiconductor End Use; Updates and Corrections," *Federal Registry* 0694-AI94, October 25, 2023. https://www.federalregister.gov/d/2023-23055

[8] "G7 Leaders' Statement on the Hiroshima AI Process," Ministry of Foreign Affairs of Japan, accessed October 28, 2024, https://www.mofa.go.jp/ecm/ec/page5e_000076.html.

[9] State Council, "Notice of the State Council on Issuing 'Made in China 2025'," Central People's Government of the People's Republic of China, May 2015, https://www.gov.cn/zhengce/content/2015-05/19/content_9784.htm

[10] Xia-heng Zhang, "Political and Social Dynamics, Risks, and Prevention of ChatGPT," *Journal of Shenzhen University (Humanities & Social Sciences)* 40, no. 3 (2023): 8.

[11] Sun Ninghui, "The Development of Artificial Intelligence and Intelligent Computing," Lecture Notes for the Tenth Special Lecture of the 14th National People's Congress Standing Committee, April 2024, http://www.npc.gov.cn/npc/c2/c30834/202404/t20240430_436915.html

[12] Juan Luis Manfredi-Sánchez and Pablo Sebastian Morales, "Generative AI and the Future for China's Diplomacy," *Place Branding and Public Diplomacy*, March 11, 2024: 2.

[13] Robert Booth and Dan Milmo, "Chinese AI Chatbot DeepSeek Censors Itself in Realtime, Users Report," *The Guardian*, January 28, 2025, sec. Technology, https://www.theguardian.com/technology/2025/jan/28/chinese-ai-chatbot-deepseek-censors-itself-in-realtime-users-report.

[14] Google Ngram Viewer, "话语权," https://goo.gl/sUJ8la

[15] Lu Wei, "National Discourse Power and Information Security Against the Background of Economic Globalization," trans. Rogier Creemers, DigiChina - Stanford Cyber Policy Center, December 21, 2014, https://digichina.stanford.edu/work/national-discourse-power-and-information-security-against-the-background-of-economic-globalization/.

[16] China Academy of Information and Communications Technology. "Cloud Computing White Paper," *CAICT*, September 2016. https://www.cac.gov.cn/files/pdf/baipishu/cloudcomputing2016.pdf

[17] Chen Jiancai, "Theoretical Communication Needs to Have a Strong Trunk and a Flourishing Crown," *网络传播 (New Media)*, February 2016: 48. https://www.cac.gov.cn/files/pdf/zazhiwlcb/zazhiwlcb201602.pdf [emphasis added]

[18] Mirzoeff, *The Right to Look*: 1.

[19] "China's Latest AI Chatbot Is Trained on President Xi Jinping's Political Ideology," AP News, May 24, 2024, https://apnews.com/article/china-chatbot-xi-jinping-thought-ai-chatgpt-68e6d58031f2fc55dc3c569dd1f0afb2. ; Xi Jinping, "Secure a Decisive Victory in Building a Moderately Prosperous Society in All Respects and Strive for the Great Success of Socialism with Chinese Characteristics for a New Era," transcript of speech delivered at the 19th National Congress of the Communist Party of China, Beijing, October 18, 2017, http://www.xinhuanet.com/english/download/Xi_Jinping's_report_at_19th_CPC_National_Congress.pdf

[20] Michael Caster, "Apple-Baidu Partnership Risks Accelerating China's Influence Over the Future of Generative AI," The Diplomat, April 17, 2024, https://thediplomat.com/2024/04/apple-baidu-partnership-risks-accelerating-chinas-influence-over-the-future-of-generative-ai/.

[21] Thomas Eder, Rebecca Arcesati, and Jacob Mardell, "Networking the 'Belt and Road' - The Future Is Digital," Mercator Institute for China Studies, August 28, 2019, https://merics.org/en/tracker/networking-belt-and-road-future-digital.

[22] Zongyuan Zoe Liu, "China's Real Economic Crisis," *Foreign Affairs*, August 6, 2024, https://www.foreignaffairs.com/china/chinas-real-economic-crisis-zongyuan-liu.

[23] Jacob Funk Kirkegaard, "China's Population Decline Is Getting Close to Irreversible," Peterson Institute for International Economics, January 18, 2024, https://www.piie.com/research/piie-charts/2024/chinas-population-decline-getting-close-irreversible.

[24] "China Dissent Monitor - Issue 8: April – June 2024," Freedom House, accessed October 31, 2024, https://freedomhouse.org/report/china-dissent-monitor/2024/issue-8-april-june-2024.

[25] Katja Drinhausen and Helena Legarda, "'Comprehensive National Security' Unleashed: How Xi's Approach Shapes China's Policies at Home and Abroad," MERICS China Monitor (Berlin: Mercator Institute for China Studies, September 15, 2022), https://merics.org/en/report/comprehensive-national-security-unleashed-how-xis-approach-shapes-chinas-policies-home-and.

[26] Tony Roberts et al., "Mapping the Supply of Surveillance Technologies to Africa: Case Studies from Nigeria, Ghana, Morocco, Malawi, and Zambia" (Brighton: Institute of Development Studies, September 27, 2023): 5.

[27] Claudio Feijóo et al., "Harnessing Artificial Intelligence (AI) to Increase Wellbeing for All: The Case for a New Technology Diplomacy," *Telecommunications Policy*, Artificial intelligence, economy and society, 44, no. 6 (July 1, 2020): 10. https://doi.org/10.1016/j.telpol.2020.101988.

[28] "America's AI Action Plan" (Executive Office of the President of the United States, July 2025): 1.

[29] "Executive Order 14320 of July 23,2025, Promoting the Export of the American AI Technology Stack," *Code of Federal Regulations*, 2025-14218 (2025): 35393, https://www.federalregister.gov/d/2025-14218

[30] Ibid., 35397.

[31] Zoe Kleinman and Liv McMahon, "UK and US Refuse to Sign International AI Declaration," BBC News, February 11, 2025, https://www.bbc.com/news/articles/c8edn0n58gwo.

[32] Ministry of Foreign Affairs, *Global AI Governance Action Plan*, July 26, 2025, https://www.mfa.gov.cn/zyxw/202507/t20250726_11677803.shtml

[33] Mirzoeff, *The Right to Look*, 23.

[34] State Internet Information Office, *Interim Measures for Generative Artificial Intelligence Service Management*, July 2023: Article 4(1). https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

[35] Barry Friedman, "What Is Public Safety?," *Boston University Law Review* 102, no. 3 (2022): 739.

CONCLUSION

[1] Mac History, "Apple Knowledge Navigator Video (1987)," Youtube, March 4, 2012, video, 05:45, https://www.youtube.com/watch?v=umJsITGzXd0.

[2] Nicholas Negroponte, *Being Digital* (London: Hodder & Stoughton, 1995): 151.

[3] Ibid., 153.

[4] Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (Penguin Press, 2011): 18.

[5] "Claude's Constitution," Anthropic, May 9, 2023, https://www.anthropic.com/news/claudes-constitution.

[6] "Project Exo," Channel 1 AI, accessed May 14, 2025, https://www.channel1.ai/exo.

[7] Neal Stephenson, *The Diamond Age* (New York: Bantam Books, 1995): 74.

[8] Stephenson, *The Diamond Age*: 40.

[9] Stephenson, *The Diamond Age*: 83.

[10] Ibid., 452.

[11] Mirzoeff, *The Right to Look*: 34.

[12] Ibid.

[13] Hans Magnus Enzensberger, "Constituents of a Theory of the Media," *New Left Review* 0, no. 64 (November 1, 1970): 30.

[14] Mirzoeff, *The Right to Look*: 3.

[15] Bernard Stiegler, *For a New Critique of Political Economy*, 1st ed. (Polity, 2010): 123.

[16] Ibid.

[17] Louise Amoore et al., "A World Model: On the Political Logics of Generative AI," *Political Geography* 113 (August 1, 2024): 2.

[18] Ibid., 8.

[19] José Esteban Muñoz, *Cruising Utopia: The Then and There of Queer Futurity* (NYU Press, 2009): 23.

[20] Phil Jones, "Make It New," *Verso* (blog), June 20, 2023, https://www.versobooks.com/en-ca/blogs/news/make-it-new.

[21] Megan Hyska, "The Politics of Past and Future: Synthetic Media, Showing, and Telling," *Philosophical Studies*, November 21, 2023: 4.

[22] Ziv Epstein, Hope Schroeder, and Dava Newman, "When Happy Accidents Spark Creativity: Bringing Collaborative Speculation to Life with Generative AI" (arXiv, June 17, 2022), https://doi.org/10.48550/arXiv.2206.00533.

[23] Hyska, "The Politics of Past and Future […],": 18.

[24] Jonas Staal, *Propaganda Art in the 21st Century* (MIT Press, 2019): 59.

[25] Ibid., 3.

[26] Ibid., 1.

[27] Ibid., 50.

[28] Enzensberger, "Constituents of a Theory of the Media,": 21.

[29] Ibid., 20.

[30] Ahmed M. Abuzuraiq and Philippe Pasquier, "Seizing the Means of Production: Exploring the Landscape of Crafting, Adapting and Navigating Generative AI Models in the Visual Arts" (arXiv, April 26, 2024), https://doi.org/10.48550/arXiv.2404.17688.

[31] Louise Amoore, "The Deep Border," *Political Geography* 109 (March 1, 2024): 8.

[32] Jonas Staal, "Assemblism - Journal #80," e-flux, accessed May 23, 2025, https://www.e-flux.com/journal/80/100465/assemblism/.

[33] Staal, *Propaganda Art* […]: 189.

[34] Foucault, *Discipline and Punish*: 82.

[35] A. J. Bauer and Anthony Nadler, "What Is Disinformation to Democracy?," *Center for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill*, January 23, 2023: 5.

[36] Mark Fisher, *Capitalist Realism: Is There No Alternative?* (Zero Books, 2009): 80.

BIBLIOGRAPHY

Abrahams, Alexei, and Gabrielle Lim. “Repress/Redress: What the ‘War on Terror’ Can Teach Us about Fighting Misinformation.” *Harvard Kennedy School Misinformation Review*, July 22, 2020. https://doi.org/10.37016/mr-2020-032.

Abuzuraiq, Ahmed M., and Philippe Pasquier. “Seizing the Means of Production: Exploring the Landscape of Crafting, Adapting and Navigating Generative AI Models in the Visual Arts.” arXiv, April 26, 2024. https://doi.org/10.48550/arXiv.2404.17688.

Acerbi, Alberto, Sacha Altay, and Hugo Mercier. “Research Note: Fighting Misinformation or Fighting for Information?” *Harvard Kennedy School Misinformation Review*, January 12, 2022. https://doi.org/10.37016/mr-2020-87.

———. “Research Note: Fighting Misinformation or Fighting for Information?” *Harvard Kennedy School Misinformation Review*, January 12, 2022. https://doi.org/10.37016/mr-2020-87.

Adorno, Theodor W. and Max Horkheimer. “The Culture Industry: Enlightenment as Mass Deception.” In *Dialectic of Enlightenment: Philosophical Fragments*. Edited by Gunzelin Schmid Noerr and translated by Edmund Jephcott. Stanford, CA: University of Stanford Press, 2002: 94–136

Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami. “Mining Association Rules between Sets of Items in Large Databases.” In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 207–16. Washington D.C. USA: ACM, 1993. https://doi.org/10.1145/170035.170072.

Akers, John, Gagan Bansal, Gabriel Cadamuro, Christine Chen, Quanze Chen, Lucy Lin, Phoebe Mulcaire, et al. “Technology-Enabled Disinformation: Summary, Lessons, and Recommendations.” arXiv, January 3, 2019. https://doi.org/10.48550/arXiv.1812.09383.

“AIML-TUDA/I2p,” Datasets at Hugging Face, accessed April 8, 2024, https://huggingface.co/datasets/AIML-TUDA/i2p.

Alam, Firoj, Stefano Cresci, Tanmoy Chakraborty, Fabrizio Silvestri, Dimiter Dimitrov, Giovanni Da San Martino, Shaden Shaar, Hamed Firooz, and Preslav Nakov. “A Survey on Multimodal Disinformation Detection.” arXiv, September 28, 2022. https://doi.org/10.48550/arXiv.2103.12541.

Aldane, Jack. “Tech Firms Sign Pact to Counter AI Deepfakes as Elections near - Global Government Forum.” Global Government Forum, February 19, 2024. https://www.globalgovernmentforum.com/tech-firms-sign-pact-to-counter-ai-deepfakes-as-elections-near/.

Allen, Danielle, and E. Glen Weyl. “The Real Dangers of Generative AI.” *Journal of Democracy* 35, no. 1 (2024): 147–62.

Allen, Jennifer, Baird Howland, Markus Mobius, David Rothschild, and Duncan J. Watts. “Evaluating the Fake News Problem at the Scale of the Information Ecosystem.” *Science Advances* 6, no. 14 (April 3, 2020): 3539. https://doi.org/10.1126/sciadv.aay3539.

Altay, Sacha, Manon Berriche, and Alberto Acerbi. “Misinformation on Misinformation: Conceptual and Methodological Challenges.” *Social Media + Society* 9, no. 1 (January 1, 2023): 20563051221150412. https://doi.org/10.1177/20563051221150412.

Altay, Sacha, Rasmus Kleis Nielsen, and Richard Fletcher. “Quantifying the ‘Infodemic’: People Turned to Trustworthy News Outlets during the 2020 Coronavirus Pandemic.” *Journal of Quantitative Description: Digital Media* 2 (August 26, 2022). https://doi.org/10.51685/jqd.2022.020.

Amarikwa, Melany. “Internet Openness at Risk: Generative AI’s Impact on Data Scraping.” SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, February 5, 2024. https://doi.org/10.2139/ssrn.4723713.

Amoore, Louise. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham: Duke University Press, 2020.

———. “Lines of Sight: On the Visualization of Unknown Futures.” *Citizenship Studies* 13, no. 1 (February 1, 2009): 17–30. https://doi.org/10.1080/13621020802586628.

———. “The Deep Border.” *Political Geography* 109 (March 1, 2024): 102547. https://doi.org/10.1016/j.polgeo.2021.102547.

Amoore, Louise, Alexander Campolo, Benjamin Jacobsen, and Ludovico Rella. “A World Model: On the Political Logics of Generative AI.” *Political Geography* 113 (August 1, 2024): 103134. https://doi.org/10.1016/j.polgeo.2024.103134.

Andrejevic, Mark. *Automated Media*. London ; New York, NY: Routledge, 2020.

Anthropic. “Anthropic Awarded $200M DOD Agreement for AI Capabilities,” July 14, 2025, https://www.anthropic.com/news/anthropic-and-the-department-of-defense-to-advance-responsible-ai-in-

defense-operations.
———. “Claude’s Constitution,” May 9, 2023. https://www.anthropic.com/news/claudes-constitution.
———. “Model Card and Evaluations for Claude Models,” July 8, 2023. https://wwwcdn.anthropic.com/bd2a28d2535bfb0494cc8e2a3bf135d2e7523226/Model-Card-Claude-2.pdf
———. “Preparing for Global Elections in 2024,” February 16, 2024. https://www.anthropic.com/news/preparing-for-global-elections-in-2024.
AP News. “China’s Latest AI Chatbot Is Trained on President Xi Jinping’s Political Ideology,” May 24, 2024. https://apnews.com/article/china-chatbot-xi-jinping-thought-ai-chatgpt-68e6d58031f2fc55dc3c569dd1f0afb2.
AP News. “Insider Q&A: CIA’s Chief Technologist’s Cautious Embrace of Generative AI,” May 20, 2024. https://apnews.com/article/cia-chief-technologist-gen-ai-c3f4154a8e2879d7820931f07656968c.
Apple Newsroom (Canada). “Apple Unveils M3, M3 Pro, and M3 Max, the Most Advanced Chips for a Personal Computer.” Accessed February 13, 2024. https://www.apple.com/ca/newsroom/2023/10/apple-unveils-m3-m3-pro-and-m3-max-the-most-advanced-chips-for-a-personal-computer/.
Apprich, Clemens, Wendy Hui Kyong Chun, Florian Cramer, and Hito Steyerl. *Pattern Discrimination*. Paperback. Meson Press, University of Minnesota Press, 2018.
Aristotle, *Nicomachean Ethics*, ed. Roger Crisp (Cambridge University Press, 2000).
Askonas, Jon. “What Happened to Consensus Reality?” *The New Atlantis*, no. Spring 2022 (June 30, 2022). https://www.thenewatlantis.com/publications/what-happened-to-consensus-reality.
Aspen Digital. “AI’s Impact on 2024 Global Elections,” March 28, 2024. https://www.aspendigital.org/event/ai-2024-global-elections/.
Atanasoski, Neda, and Kalindi Vora. *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*. Duke University Press, 2019.
Aythora, J., R. Burke-Aguero, A. Chamayou, and S. Clebsch. “MULTI-STAKEHOLDER MEDIA PROVENANCE MANAGEMENT TO COUNTER SYNTHETIC MEDIA RISKS IN NEWS PUBLISHING.” virtual, 2020. https://drive.google.com/file/d/11c41iTwq7z-nMMMPQLE5GZ6gJmc_llJ1/view?usp=sharing&usp=embed_facebook.
Azoulay, Pierre, Joshua L. Krieger, and Abhishek Nagaraj. “Old Moats for New Models: Openness, Control, and Competition in Generative AI.” Working Paper. Working Paper Series. National Bureau of Economic Research, May 2024. https://doi.org/10.3386/w32474.
Bai, Yuntao, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, et al. “Constitutional AI: Harmlessness from AI Feedback.” arXiv, December 15, 2022. https://doi.org/10.48550/arXiv.2212.08073.
Bajohr, Hannes. “Whoever Controls Language Models Controls Politics.” *Hannes Bajohr* (blog), April 8, 2023. https://hannesbajohr.de/en/2023/04/08/whoever-controls-language-models-controls-politics/.
Baldridge, David, Beth Coleman, and Jamie Sandhu. “The Terminology of AI Regulation: Preventing ‘Harm’ and Mitigating ‘Risk.’” Schwartz Reisman Institute. Accessed March 18, 2024. https://srinstitute.utoronto.ca/news/terminology-regulation-risk-harm.
Balzacq, Thierry. “The Three Faces of Securitization: Political Agency, Audience and Context.” *European Journal of International Relations* 11, no. 2 (June 2005): 171–201. https://doi.org/10.1177/1354066105052960.
Bardin, Andrea, and Marco Ferrari. “Governing Progress: From Cybernetic Homeostasis to Simondon’s Politics of Metastability.” *The Sociological Review* 70, no. 2 (March 1, 2022): 248–63. https://doi.org/10.1177/00380261221084426.
Barker, Hannah. *Newspapers, Politics and English Society, 1695-1855*. Longman, 2000.
Barman, Dipto, Ziyi Guo, and Owen Conlan. “The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination.” *Machine Learning with Applications* 16 (June 1, 2024): 100545. https://doi.org/10.1016/j.mlwa.2024.100545.
Basta, Christine, Marta R. Costa-jussà, and Noe Casas. “Evaluating the Underlying Gender Bias in Contextualized Word Embeddings.” arXiv, April 18, 2019. https://doi.org/10.48550/arXiv.1904.08783.
Bastian, Matthias. “DALL-E 3’s System Prompt Reveals OpenAI’s Rules for AI Image Generation.” THE DECODER, October 16, 2023. https://the-decoder.com/dall-e-3s-system-prompt-reveals-openais-rules-for-generative-image-ai/.
Bastos, Marco T., and Dan Mercea. “The Brexit Botnet and User-Generated Hyperpartisan News.” *Social Science Computer Review* 37, no. 1 (February 1, 2019): 38–54. https://doi.org/10.1177/0894439317734157.
Baudrillard, Jean. “The Precession of Simulacra” in *Simulacra and Simulation* (Ann Arbor: University of Michigan Press, 1994).

Bauer, A. J., and Anthony Nadler. “What Is Disinformation to Democracy?” *Center for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill*, January 23, 2023. https://citap.pubpub.org/pub/u4qzry5p/release/1.

Bauer, Fred. “Why We Must Resist AI’s Soft Mind Control.” *The Atlantic* (blog), March 8, 2024. https://www.theatlantic.com/ideas/archive/2024/03/artificial-intelligence-google-gemini-mind-control/677683/.

BBC. “Google to Fix AI Picture Bot after ‘woke’ Criticism,” February 21, 2024. https://www.bbc.com/news/business-68364690.

Beck, Ulrich. *Risk Society: Towards a New Modernity*. SAGE, 1992.

Bellemare, Andrea, and Jason Ho. “Social Media Firms Catching More Misinformation, but Critics Say They Could Be Doing More,” *CBC News*, April 20, 2020. https://www.cbc.ca/news/science/social-media-platforms-pandemic-moderation-1.5536594.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜.” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. FAccT ’21. New York, NY, USA: Association for Computing Machinery, 2021. https://doi.org/10.1145/3442188.3445922.

Bengio, Yoshua, Eric Laufer, Guillaume Alain, and Jason Yosinski. “Deep Generative Stochastic Networks Trainable by Backprop.” In *Proceedings of the 31st International Conference on Machine Learning*, 226–34. PMLR, 2014. https://proceedings.mlr.press/v32/bengio14.html.

Benjamin, Garfield. “What We Do with Data: A Performative Critique of Data ‘Collection.’” *Internet Policy Review* 10, no. 4 (2021): 1–27. https://doi.org/10.14763/2021.4.1588.

Benjamin, Ruha. *Race after Technology: Abolitionist Tools for the New Jim Code*. Cambridge, England ; Polity Press, 2019.

Benkler, Yochai. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press, 2008. https://doi.org/10.12987/9780300127232.

Benkler, Yochai, Robert Faris, and Hal Roberts. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press, 2018.

Bennett, W. Lance, and Steven Livingston. “A Brief History of the Disinformation Age: Information Wars and the Decline of Institutional Authority.” In *The Disinformation Age*, edited by Steven Livingston and W. Lance Bennett, 3–40. SSRC Anxieties of Democracy. Cambridge: Cambridge University Press, 2020. https://doi.org/10.1017/9781108914628.001.

Berger, J. M. “Our Consensus Reality Has Shattered.” *The Atlantic* (blog), October 9, 2020. https://www.theatlantic.com/ideas/archive/2020/10/year-living-uncertainly/616648/.

Berghel, Hal. “Generative Artificial Intelligence, Semantic Entropy, and the Big Sort.” *Computer* 57, no. 1 (January 2024): 130–35. https://doi.org/10.1109/MC.2023.3331594.

Berlinski, Elise, Jérémy Morales, and Samuel Sponem. “Artificial Imaginaries: Generative AIs as an Advanced Form of Capitalism.” *Critical Perspectives on Accounting* 99 (March 1, 2024): 102723. https://doi.org/10.1016/j.cpa.2024.102723.

Bernays, Edward L. *Propaganda: The Public Mind in the Making*. Liveright, 1936.

Bernstein, Joseph. “Bad News: Selling the Story of Disinformation.” *Harper’s Magazine*. Accessed March 18, 2024. https://harpers.org/archive/2021/09/bad-news-selling-the-story-of-disinformation/.

Berry, David M. “The Poverty of Networks.” *Theory, Culture & Society* 25, no. 7–8 (December 1, 2008): 364–72. https://doi.org/10.1177/0263276408097813.

Berry, David M. “The Relevance of Understanding Code to International Political Economy.” *International Politics* 49, no. 2 (March 1, 2012): 277–96. https://doi.org/10.1057/ip.2011.37.

Berry, David M., and James Stockman. “Schumacher in the Age of Generative AI: Towards a New Critique of Technology.” *European Journal of Social Theory*, March 1, 2024, 13684310241234028. https://doi.org/10.1177/13684310241234028.

Betker, James, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, et al. “Improving Image Generation with Better Captions,” n.d.

Bickert, Monika. “Our Approach to Labeling AI-Generated Content and Manipulated Media.” *Meta* (blog), April 5, 2024. https://about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/.

Biddle, Sam. “OpenAI Quietly Deletes Ban on Using ChatGPT for ‘Military and Warfare.’” *The Intercept* (blog), January 12, 2024. https://theintercept.com/2024/01/12/open-ai-military-ban-chatgpt/.

Bittman, Ladislav. *The KGB and Soviet Disinformation: An Insider’s View*. Pergamon-Brassey’s, 1985.

Bletchley Park. “Bletchley Park Makes History Again as Host of the World’s First AI Safety Summit,” November 1, 2023. https://bletchleypark.org.uk/bletchley-park-makes-history-again-as-host-of-the-worlds-first-ai-safety-summit/.

Bleyer-Simon, Konrad. “Government Repression Disguised as Anti-Disinformation Action: Digital Journalists’ Perception of COVID-19 Policies in Hungary.” *Journal of Digital Media & Policy* 12, no. COVID-19 and Digital Media Policy (March 1, 2021): 159–76. https://doi.org/10.1386/jdmp_00053_1.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al. “On the Opportunities and Risks of Foundation Models.” arXiv, July 12, 2022. https://doi.org/10.48550/arXiv.2108.07258.

Bontridder, Noémi, and Yves Poullet. “The Role of Artificial Intelligence in Disinformation.” *Data & Policy* 3 (January 2021): e32. https://doi.org/10.1017/dap.2021.20.

Booth, Robert, and Dan Milmo. “Chinese AI Chatbot DeepSeek Censors Itself in Realtime, Users Report.” *The Guardian*, January 28, 2025, sec. Technology. https://www.theguardian.com/technology/2025/jan/28/chinese-ai-chatbot-deepseek-censors-itself-in-realtime-users-report.

Bouey, Jennifer, Lynn Hu, Keller Scholl, William Marcellino, Rafiq Dossani, Ammar A. Malik, Kyra Solomon, Sheng Zhang, and Andy Shufer. “China’s AI Exports: Technology Distribution and Data Safety.” RAND Corporation, December 11, 2023. https://www.rand.org/pubs/research_reports/RRA2696-2.html.

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. Inside Technology. MIT Press, 1999.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. “A Note on Internet Use and the 2016 U.S. Presidential Election Outcome.” *PLOS ONE* 13, no. 7 (July 18, 2018): e0199571. https://doi.org/10.1371/journal.pone.0199571.

boyd, danah, and Kate Crawford. “Critical Questions for Big Data.” *Information, Communication & Society* 15, no. 5 (June 2012): 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Bratich, Jack. “Civil Society Must Be Defended: Misinformation, Moral Panics, and Wars of Restoration.” *Communication, Culture and Critique* 13, no. 3 (September 1, 2020): 311–32. https://doi.org/10.1093/ccc/tcz041.

Bratich, Jack Z. *Conspiracy Panics: Political Rationality and Popular Culture*. Albany: SUNY Press, 2008.

Brayne, Sarah. *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford University Press, 2021.

Bresnick, Sam. “China Bets Big on Military AI.” CEPA, April 3, 2024. https://cepa.org/article/china-bets-big-on-military-ai/.

Brittain, Blake. “How Copyright Law Could Threaten the AI Industry in 2024.” Reuters. Accessed March 21, 2024. https://www.reuters.com/legal/litigation/how-copyright-law-could-threaten-ai-industry-2024-2024-01-02/.

Brodsky, Sascha. “Tech Industry Ramps up Efforts to Combat Rising Deepfake Threats.” *IBM Blog* (blog), September 10, 2024. https://www.ibm.com/blog/deepfake-detection/.

Broinowski, Anna. “Deepfake Nightmares, Synthetic Dreams: A Review of Dystopian and Utopian Discourses Around Deepfakes, and Why the Collapse of Reality May Not Be Imminent—Yet.” *Journal of Asia-Pacific Pop Culture* 7, no. 1 (2022): 109–39.

Broussard, and Meredith. *Artificial Unintelligence: How Computers Misunderstand the World*. Artificial Unintelligence. MIT Press Ltd, 2018.

Broussard, Meredith. *More than a Glitch: Confronting Race, Gender, and Ability Bias in Tech*. MIT Press, 2023.

Brunila, Mikael. “Taking AI into the Tunnels.” e-flux, February 2025. https://www.e-flux.com/journal/151/652643/taking-ai-into-the-tunnels/.

Bucher, Taina. *If...Then: Algorithmic Power and Politics*. Oxford Studies in Digital Politics. Oxford University Press, USA, 2018.

———. “Want to Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook.” *New Media & Society* 14, no. 7 (2012): 1164–80. https://doi.org/10.1177/1461444812440159.

Buel, Jason W. “Automated Visions, Algorithmic Imageflows: The Technopolitics of Black Lives Matter Videos on YouTube.” *MAST The Journal of Media Art Study and Theory* 3, no. 1 (April 2022): 113–33.

Bui, Anh, Long Vuong, Khanh Doan, Trung Le, Paul Montague, Tamas Abraham, and Dinh Phung. “Erasing Undesirable Concepts in Diffusion Models with Adversarial Preservation.” arXiv, October 29, 2024. https://doi.org/10.48550/arXiv.2410.15618.

Buolamwini, Joy, and Timnit Gebru. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. PMLR, 2018. https://proceedings.mlr.press/v81/buolamwini18a.html.

Burgess, Matt. “How to Stop Your Data From Being Used to Train AI.” *Wired*. Accessed March 14, 2025. https://www.wired.com/story/how-to-stop-your-data-from-being-used-to-train-ai/.

Burke, Garance. “Researchers Say an AI-Powered Transcription Tool Used in Hospitals Invents Things No One Ever Said.” AP News, October 26, 2024. https://apnews.com/article/ai-artificial-intelligence-health-business-90020cdf5fa16c79ca2e5b6c4c9bbb14.

Burns, A., J. McDermid, and J. Dobson. “On the Meaning of Safety and Security.” *The Computer Journal* 35, no. 1 (February 1, 1992): 3–15. https://doi.org/10.1093/comjnl/35.1.3.

Burroughs, William S. *The Electronic Revolution*. (Cambridge: Blackmoor Head Press, 1971)

———. “The Limits of Control.” *Semiotext(e): Schizo-Culture* 3.2 (1978).

Buzan, Barry, Ole Wæver, and Jaap de Wilde. *Security: A New Framework for Analysis*. Lynne Rienner Publishers, 2022.

Cadwalladr, Carole, and Emma Graham-Harrison. “Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach.” *The Guardian*, March 17, 2018, sec. News. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Calabrese, Orsolya Reich, Sofia. “Beyond Disinformation: How DSA Risk Assessments Ignore Democracy’s Real Threats.” Policy Analysis. Tech Policy Press, March 19, 2025. https://techpolicy.press/beyond-disinformation-how-dsa-risk-assessments-ignore-democracys-real-threats.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. “Semantics Derived Automatically from Language Corpora Contain Human-like Biases.” *Science* 356, no. 6334 (April 14, 2017): 183–86. https://doi.org/10.1126/science.aal4230.

Call, Christchurch. “Four New Tech Firms Expand the Christchurch Call.” Christchurch Call. Accessed May 28, 2024. https://www.christchurchcall.com/media-and-resources/news-and-updates/four-new-tech-firms-join-the-christchurch-call.

———. “Understanding Algorithms and Developing Interventions.” Christchurch Call. Accessed May 29, 2024. https://www.christchurchcall.com/our-work/algorithms-and-positive-interventions.

Calvet-Bademunt, Jordi, and Jacob Mchangama. “Freedom of Expression in Generative AI: A Snapshot of Content Policies.” The Future of Free Speech, February 2024.

Carlyle, Thomas, Sarah Atwood, Owen Dudley Edwards, Christopher Harvie, Terence James Reed, and Beverly Taylor. *On Heroes, Hero Worship, and the Heroic in History*. Yale University Press, 2013. https://www.jstor.org/stable/j.ctt5vm0w4.

Cassirer, Ernst. *An Essay on Man: An Introduction to a Philosophy of Human Culture*. New Haven, CT: Yale University Press, 1992.

Caster, Michael. “Apple-Baidu Partnership Risks Accelerating China’s Influence Over the Future of Generative AI.” The Diplomat, April 17, 2024. https://thediplomat.com/2024/04/apple-baidu-partnership-risks-accelerating-chinas-influence-over-the-future-of-generative-ai/.

Castro, Clinton, Pham, and Alan Rubel. “Epistemic Paternalism Online.” In *Epistemic Paternalism: Conceptions, Justifications and Implications*, edited by Guy Axtell and Amiel Bernal. Rowman & Littlefield, 2020.

Central Committee of the Communist Party of China, “Encouraging ground-breaking scientific and technological innovation” in *Outline of the 14th Five-Year Plan (2021-2025) for National Economic and Social Development and Vision 2035 of the People's Republic of China*, https://www.fujian.gov.cn/english/news/202108/t20210809_5665713.htm

———. “New Information Network Technology” in *The 13th Five-Year Plan for Economic and Social Development of the People’s Republic of China 2016-2020*, https://en.ndrc.gov.cn/policies/202105/P020210527785800103339.pdf

Central Intelligence Agency. “A HISTORY OF THE HAMLET EVALUATION SYSTEM (HES) | CIA FOIA (Foia.Cia.Gov).” Accessed August 31, 2023. https://www.cia.gov/readingroom/document/cia-rdp80r01720r000200120001-2.

Chakrabarty, Dipesh. *Provincializing Europe: Postcolonial Thought and Historical Difference*. Princeton Studies in Culture / Power / History. Princeton (N.J.): Princeton university press, 2007.

Channel 1 AI. “Project Exo.” Accessed May 14, 2025. https://www.channel1.ai/exo.

Chen, Adrian. “The Agency.” *The New York Times Magazine*, June 2, 2015. https://www.nytimes.com/2015/06/07/magazine/the-agency.html.

Cheng, Yang, and Zifei Fay Chen. “The Influence of Presumed Fake News Influence: Examining Public Support for Corporate Corrective Response, Media Literacy Interventions, and Governmental Regulation.” *Mass Communication and Society* 23, no. 5 (September 2, 2020): 705–29. https://doi.org/10.1080/15205436.2020.1750656.

Chesney, Robert, and Danielle Keats Citron. “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security.” *SSRN Electronic Journal*, 2018, 1753–1820. https://doi.org/10.2139/ssrn.3213954.

Chiang, Ted. “ChatGPT Is a Blurry JPEG of the Web.” *The New Yorker*, February 9, 2023. https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web.

China Academy of Information and Communications Technology. “Cloud Computing White Paper,” *CAICT*, September 2016. https://www.cac.gov.cn/files/pdf/baipishu/cloudcomputing2016.pdf

Chinese National Information Security Standardization Technical Committee. *Technical Documentation of the National Information Security Standardization Technical Committee: Basic Safety Requirements for Generative Artificial Intelligence Services (Draft for Feedback),* Ben Murphy trans., Oct. 2023, https://cset.georgetown.edu/publication/china-safety-requirements-for-generative-ai/

Chow, Andrew R. “AI’s Underwhelming Impact On the 2024 Elections.” TIME, October 30, 2024. https://time.com/7131271/ai-2024-elections/.

Christchurch Call Foundation. “Christchurch Call Leaders’ Summit 2023: Compilation of supporting papers,” Accessed May 27, 2024. https://www.christchurchcall.com/assets/Documents/Christchurch-Call-Leaders-Summit-2023Supporting-papers.pdf

———. “Christchurch Call Text.” Accessed May 28, 2024. https://www.christchurchcall.com/about/christchurch-call-text.

Chui, Michael, Martin Harryson, James Manyika, and Roger Roberts. “NOTES FROM THE AI FRONTIER: APPLYING AI FOR SOCIAL GOOD.” Discussion Paper. McKinsey Global Institute, December 2018. https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.pdf.

Chun, Wendy Hui Kyong, and Alex H. Barnett. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, Massachusetts London, England: The MIT Press, 2021.

Chun, Wendy Hui-Kyong. *Control and Freedom: Power and Paranoia in the Age of Fiber Optics*. Cambridge, Mass: MIT Press, 2006.

Cohen, Jared, and George Lee. “The Generative World Order: AI, Geopolitics, and Power.” Goldman Sachs Insights, December 14, 2023. https://www.goldmansachs.com/insights/articles/the-generative-world-order-ai-geopolitics-and-power.

Cohen, Stanley. *Folk Devils and Moral Panics*. New York: Routledge, 1972.

Cole, Matthew, Hugo Radice, and Charles Umney. “The Political Economy of Datafication and Work: A New Digital Taylorism?” *Socialist Register* 57 (2021). https://socialistregister.com/index.php/srv/article/view/34948.

“Common Crawl - Open Repository of Web Crawl Data.” Accessed April 2, 2024. https://commoncrawl.org/.

Content Authenticity Initiative. “Introducing the Content Authenticity Initiative,” November 4, 2019. https://contentauthenticity.org/blog/test.

Coalition for Content Provenance and Authenticity. “About,” Accessed May 5, 2024. https://c2pa.org/about/about/

———. “C2PA Founding Press Release,” February 22, 2021. https://c2pa.org/post/c2pa_initial_pr/.

———. *C2PA Harms Modelling*, accessed May 6, 2024, https://c2pa.org/specifications/specifications/1.0/security/Harms_Modelling.html

———. *C2PA Technical Specification*, accessed April 2, 2024, https://c2pa.org/specifications/specifications/1.0/specs/C2PA_Specification.html

Cooke, Di. “Synthetic Media and Election Integrity: Defending Our Democracies.” Centre for Emerging Technology and Security, August 2023.

Cooke, Di, Abby Edwards, Alexis Day, Devi Nair, Sophia Barkoff, and Katie Kelly. “Crossing the Deepfake Rubicon: The Maturing Synthetic Media Threat Landscape.” Center for Strategic and International Studies (CSIS), 2024. https://www.jstor.org/stable/resrep64562.

Costello, Thomas H., Gordon Pennycook, and David Rand. “Durably Reducing Conspiracy Beliefs through Dialogues with AI.” OSF, April 3, 2024. https://doi.org/10.31234/osf.io/xcwdn.

Council of the European Union, *EU Artificial Intelligence Act*, 26 January 2024, 63, https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf

Crawford, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.

Crawford, Kate, and Trevor Paglen. “Excavating AI: The Politics of Images in Machine Learning Training Sets.” *AI & SOCIETY* 36, no. 4 (December 1, 2021): 1105–16. https://doi.org/10.1007/s00146-021-01162-8.

Crevier, Daniel. *Ai: The Tumultuous History Of The Search For Artificial Intelligence*. Basic Books, 1993.

Crouse, Megan. “Could C2PA Cryptography Be the Key to Fighting AI-Driven Misinformation?” TechRepublic, August 3, 2023. https://www.techrepublic.com/article/ai-cryptography-watermarking/.
D4BL. “About - Data For Black Lives.” Accessed March 14, 2025. https://d4bl.org/about.
Dahlgren, Peter. “Media, Knowledge and Trust: The Deepening Epistemic Crisis of Democracy.” *Javnost - The Public* 25, no. 1–2 (April 3, 2018): 20–27. https://doi.org/10.1080/13183222.2018.1418819.
“DARPA Announces Research Teams Selected to Semantic Forensics Program,” March 2, 2021. https://www.darpa.mil/news-events/2021-03-02.
Davies, Lizzy. “Occupy Movement: City-by-City Police Crackdowns so Far.” *The Guardian*, November 15, 2011, sec. US news. https://www.theguardian.com/world/blog/2011/nov/15/occupy-movement-police-crackdowns.
Dayan, Hilla. “Regimes of separation: Israel/Palestine and the shadow of apartheid,” in *The Power of Inclusive Exclusion: Anatomy of Israeli Rule in the Occupied Palestinian Territories*. Eds. Adi Ophir, M. Givoni, and Sārī Ḥanafī (New York: Zone Books, 2009).
Debord, Guy. *The Society of the Spectacle*. Translated by Donald Nicholson-Smith. Zone Books, 2020. https://doi.org/10.2307/j.ctv1453m69.
Deep Media, “Deep Media AI,” DeepMedia, Accessed April 20, 2024. https://deepmedia.ai/
Deepware, “Deepware | Scan and Detect Deepfake Videos,” *Deepware*. https://deepware.ai/
Deibert, Ronald J. “The Road to Digital Unfreedom: Three Painful Truths About Social Media.” *Journal of Democracy* 30, no. 1 (2019): 25–39.
Deleuze, Gilles. “Postscript on the Societies of Control.” *October* 59 (1992): 3–7.
Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. “ImageNet: A Large-Scale Hierarchical Image Database.” In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–55, 2009. https://doi.org/10.1109/CVPR.2009.5206848.
Denson, Shane. *Discorrelated Images*. Durham, NC: Duke University Press, 2020.

Devereaux, Zachary P, Régine Lecocq, Marc-André Labrie, and Bruce C Forrester. “Contemporary Media and the AI-Goldrush from a Military Perspective.” Carnegie Mellon Univeristy, 2023.
Dilanian, Ken, and Julia Ainsley. “DHS Weighing Major Changes to Fight Domestic Violent Extremism.” NBC News, March 25, 2021. https://www.nbcnews.com/politics/national-security/dhs-weighing-huge-changes-fight-domestic-violent-extremism-say-officials-n1262047.
DiResta, Renee. “The Digital Maginot Line.” *RibbonFarm* (blog), November 28, 2018. https://www.ribbonfarm.com/2018/11/28/the-digital-maginot-line/.
Donnarumma, Marco. “Against the Norm: Othering and Otherness in AI Aesthetics.” *Digital Culture & Society* 8, no. 2 (December 1, 2022): 39–66. https://doi.org/10.14361/dcs-2022-0205.
Douglass, Frederick. *The Color Line*. The North American Review, 1881. http://archive.org/details/jstor-25100970.
Drinhausen, Katja, and Helena Legarda. “‘Comprehensive National Security’ Unleashed: How Xi’s Approach Shapes China’s Policies at Home and Abroad.” MERICS China Monitor. Berlin: Mercator Institute for China Studies, September 15, 2022. https://merics.org/en/report/comprehensive-national-security-unleashed-how-xis-approach-shapes-chinas-policies-home-and.
Dubé, Frédérique. “The Paradox of the Commons? - Part 1.” Research Note. PRAXIS, February 13, 2025. https://praxis.encommun.io/en/n/GE5bwfF7kzO7c1iNGGLErpRNIto/.
Dubois, W. E. B. *The Souls of Black Folk*. Simon and Schuster, 2005.
Eady, Gregory, Tom Paskhalis, Jan Zilinsky, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker. “Exposure to the Russian Internet Research Agency Foreign Influence Campaign on Twitter in the 2016 US Election and Its Relationship to Attitudes and Voting Behavior.” *Nature Communications* 14, no. 1 (January 9, 2023). https://doi.org/10.1038/s41467-022-35576-9.
Edelman, Murray Jacob. “The Cardinal Political Role of Art,” in *From Art to Politics: How Artistic Creations Shape Political Conceptions*. (Chicago: University of Chicago Press, 2003).
Eder, Thomas, Rebecca Arcesati, and Jacob Mardell. “Networking the ‘Belt and Road’ - The Future Is Digital.” Mercator Institute for China Studies, August 28, 2019. https://merics.org/en/tracker/networking-belt-and-road-future-digital.
Edwards, Benj. “Google’s Hidden AI Diversity Prompts Lead to Outcry over Historically Inaccurate Images.” Ars Technica, February 22, 2024. https://arstechnica.com/information-technology/2024/02/googles-hidden-ai-diversity-prompts-lead-to-outcry-over-historically-inaccurate-images/.
———. “OpenAI Confirms That AI Writing Detectors Don’t Work.” Ars Technica, September 8, 2023. https://arstechnica.com/information-technology/2023/09/openai-admits-that-ai-writing-detectors-dont-

work/.
Eiras, Francisco, Aleksandar Petrov, Bertie Vidgen, Christian Schroeder de Witt, Fabio Pizzati, Katherine Elkins, Supratik Mukhopadhyay, et al. “Near to Mid-Term Risks and Opportunities of Open Source Generative AI.” arXiv, April 25, 2024. http://arxiv.org/abs/2404.17047.
Elkin-Koren, Niva, and Maayan Perel. “Separation of Functions for AI: Restraining Speech Regulation by Online Platforms.” *Lewis & Clark Law Review* 24, no. 3 (2020): 857–98.
Ellul, Jacques. *Propaganda: The Formation of Men’s Attitudes*. Knopf Doubleday Publishing Group, 1965.
Enzensberger, Hans Magnus. “Constituents of a Theory of the Media.” *New Left Review* 0, no. 64 (November 1, 1970): 13–36.
Epstein, Ziv, Hope Schroeder, and Dava Newman. “When Happy Accidents Spark Creativity: Bringing Collaborative Speculation to Life with Generative AI.” arXiv, June 17, 2022. https://doi.org/10.48550/arXiv.2206.00533.
étrangères, Ministère de l’Europe et des Affaires. “Joint Communiqué Issued by the President of the French Republic and the Prime Minister of New Zealand: Significant Global Progress Made under Christchurch Call (13 April 2021).” France Diplomacy - Ministry for Europe and Foreign Affairs. Accessed May 28, 2024. https://www.diplomatie.gouv.fr/en/french-foreign-policy/digital-diplomacy/news/article/joint-communique-issued-by-the-president-of-the-french-republic-and-the-prime.
EU Artificial Intelligence Act. “High-Level Summary of the AI Act,” February 27, 2024. https://artificialintelligenceact.eu/high-level-summary/.
Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin’s Publishing Group, 2018.
European Commission. “General-Purpose AI Code of Practice Now Available,” July 9, 2025. https://ec.europa.eu/commission/presscorner/detail/en/ip_25_1787.
Executive Office of the President of the United States. “America’s AI Action Plan,” July 2025. https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.
“Executive Order 14110 of October 20, 2023, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,” *Code of Federal Regulations*, 2023-24283 (2023): 75191-75226, https://www.federalregister.gov/d/2023-24283
Fallis, Don. “The Epistemic Threat of Deepfakes.” *Philosophy & Technology* 34, no. 4 (December 1, 2021): 623–43. https://doi.org/10.1007/s13347-020-00419-2.
Farkas, Johan, and Jannick Schou. “Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood.” *Javnost - The Public* 25, no. 3 (July 3, 2018): 298–314. https://doi.org/10.1080/13183222.2018.1463047.
———. *Post-Truth, Fake News and Democracy: Mapping the Politics of Falsehood*. Second edition. New York, NY: Routledge, 2024.
Fazi, M. Beatrice. “Beyond Human: Deep Learning, Explainability and Representation.” *Theory, Culture & Society* 38, no. 7–8 (December 2021): 55–77. https://doi.org/10.1177/0263276420966386.
Feijóo, Claudio, Youngsun Kwon, Johannes M. Bauer, Erik Bohlin, Bronwyn Howell, Rekha Jain, Petrus Potgieter, Khuong Vu, Jason Whalley, and Jun Xia. “Harnessing Artificial Intelligence (AI) to Increase Wellbeing for All: The Case for a New Technology Diplomacy.” *Telecommunications Policy*, Artificial intelligence, economy and society, 44, no. 6 (July 1, 2020): 101988. https://doi.org/10.1016/j.telpol.2020.101988.
Feng, K. J. Kevin, Nick Ritchie, Pia Blumenthal, Andy Parsons, and Amy X. Zhang. “Examining the Impact of Provenance-Enabled Media on Trust and Accuracy Perceptions.” *Proceedings of the ACM on Human-Computer Interaction* 7, no. CSCW2 (October 4, 2023): 270:1-270:42. https://doi.org/10.1145/3610061.
Feng, Shangbin, Chan Young Park, Yuhan Liu, and Yulia Tsvetkov. “From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models.” In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, 11737–62. Toronto, Canada: Association for Computational Linguistics, 2023. https://doi.org/10.18653/v1/2023.acl-long.656.
Feng, Tao, Chuanyang Jin, Jingyu Liu, Kunlun Zhu, Haoqin Tu, Zirui Cheng, Guanyu Lin, and Jiaxuan You. “How Far Are We From AGI: Are LLMs All We Need?” arXiv, November 24, 2024. https://doi.org/10.48550/arXiv.2405.10313.
Fineman, Mia, National Gallery of Art (U.S.), and Museum of Fine Arts Houston. *Faking It: Manipulated Photography Before Photoshop*. Metropolitan Museum of Art, 2012.
Fisher, Mark. *Capitalist Realism: Is There No Alternative?* Zero Books, 2009.
Fisher, Sarah A. “Something AI Should Tell You – The Case for Labelling Synthetic Content.” *Journal of Applied*

Philosophy 42, no. 1 (2025): 272–86. https://doi.org/10.1111/japp.12758.
Fishkin, Rand. "An Anonymous Source Shared Thousands of Leaked Google Search API Documents with Me; Everyone in SEO Should See Them." SparkToro, May 28, 2024. https://sparktoro.com/blog/an-anonymous-source-shared-thousands-of-leaked-google-search-api-documents-with-me-everyone-in-seo-should-see-them/.
Foucault, Michel. *Discipline and Punish: The Birth of the Prison*. New York: Vintage Books, 1995.
———. *Society Must Be Defended: Lectures at the College de France, 1975-76*. 1st ed. Picador, 2003.
———. *The History of Sexuality: An Introduction*. Pantheon Books, 1978.
———. *The Order of Things: An Archaeology of the Human Sciences*. Psychology Press, 2002.
Foucault, Michel, and Colin Gordon. *Power/Knowledge : Selected Interviews and Other Writings, 1972-1977*. 1st American ed. Pantheon Books, 1980.
Freedom House. "China Dissent Monitor - Issue 8: April – June 2024." Accessed October 31, 2024. https://freedomhouse.org/report/china-dissent-monitor/2024/issue-8-april-june-2024.
Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001.
Friedman, Barry. "What Is Public Safety?" *Boston University Law Review* 102, no. 3 (2022): 725–92.
Friedman, Batya, and Helen Nissenbaum. "Bias in Computer Systems." *ACM Trans. Inf. Syst.* 14, no. 3 (July 1, 1996): 330–47. https://doi.org/10.1145/230538.230561.
Friggeri, Adrien, Lada Adamic, Dean Eckles, and Justin Cheng. "Rumor Cascades." *Proceedings of the International AAAI Conference on Web and Social Media* 8, no. 1 (May 16, 2014): 101–10. https://doi.org/10.1609/icwsm.v8i1.14559.
Fuchs, Christian. *Social Media: A Critical Introduction*. 1st ed. SAGE Publications Ltd, 2013.
Gabriel, Saadia, Liang Lyu, James Siderius, Marzyeh Ghassemi, Jacob Andreas, and Asu Ozdaglar. "Generative AI in the Era of 'Alternative Facts.'" *An MIT Exploration of Generative AI*, March 27, 2024. https://doi.org/10.21428/e4baedd9.82175d26.
Galloway, Alexander R. *Gaming: Essays on Algorithmic Culture*. Electronic Mediations ; v. 18. Minneapolis: University of Minnesota Press, 2006.
———. *Protocol: How Control Exists after Decentralization*. Leonardo. Cambridge, Mass: MIT Press, 2004.
Ghiurău, David, and Daniela Elena Popescu. "Distinguishing Reality from AI: Approaches for Detecting Synthetic Content." *Computers* 14, no. 1 (January 2025): 1. https://doi.org/10.3390/computers14010001.
Gibney, Elizabeth. "China's Cheap, Open AI Model DeepSeek Thrills Scientists." *Nature* 638, no. 8049 (January 23, 2025): 13–14. https://doi.org/10.1038/d41586-025-00229-6.
Gilding, Simeon. "De-Risking Authoritarian AI: A Balanced Approach to Protecting Our Digital Ecosystems." Policy Brief. e Australian Strategic Policy Institute, 2023.
Gillespie, Tarleton. "Content Moderation, AI, and the Question of Scale." *Big Data & Society* 7, no. 2 (July 1, 2020): 2053951720943234. https://doi.org/10.1177/2053951720943234.
———. "The Relevance of Algorithms." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot. The MIT Press, 2014. https://www.researchgate.net/publication/281562384_The_Relevance_of_Algorithms.
GitHub. "List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words/En at Master · LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words · GitHub." Accessed April 2, 2024. https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words/blob/master/en.
Gomes, Ben. "Our Latest Quality Improvements for Search." Google, April 25, 2017. https://blog.google/products/search/our-latest-quality-improvements-search/.
Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep Learning*. Cambridge, Mass: The MIT Press, 2016.
Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative Adversarial Networks." arXiv, June 10, 2014. https://doi.org/10.48550/arXiv.1406.2661.
Google Ngram Viewer, "话语权," https://goo.gl/sUJ8la
Gorwa, Robert, Reuben Binns, and Christian Katzenbach. "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance." *Big Data & Society* 7, no. 1 (January 1, 2020): 2053951719897945. https://doi.org/10.1177/2053951719897945.
"GPT-4." OpenAI. Accessed February 24, 2024. https://openai.com/index/gpt-4/.
Gracia, Shanay. "Americans in Both Parties Are Concerned over the Impact of AI on the 2024 Presidential Campaign." *Pew Research Center* (blog), September 19, 2024. https://www.pewresearch.org/short-

reads/2024/09/19/concern-over-the-impact-of-ai-on-2024-presidential-campaign/.
Grant, Nico. “Google Chatbot’s A.I. Images Put People of Color in Nazi-Era Uniforms.” *The New York Times*, February 22, 2024, sec. Technology. https://www.nytimes.com/2024/02/22/technology/google-gemini-german-uniforms.html.
Gray, Mary L., and Siddharth Suri. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. HarperCollins, 2019.
Gray, Myra. “Defending the US with Biometrics.” *Infosecurity Today* 6, no. 6 (2009): 24–25. https://doi.org/10.1016/S1742-6847(09)70015-1.
Green, Ben. “The Flaws of Policies Requiring Human Oversight of Government Algorithms.” *Computer Law & Security Review* 45 (July 1, 2022): 105681. https://doi.org/10.1016/j.clsr.2022.105681.
Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. “Fake News on Twitter during the 2016 U.S. Presidential Election.” *Science* 363, no. 6425 (January 25, 2019): 374–78. https://doi.org/10.1126/science.aau2706.
Grønneberg, Hanna L., and Ana Alacovska. “Critical Dataset and Machine Learning Art.” *Morals & Machines* 2, no. 2 (2022): 22–31. https://doi.org/10.5771/2747-5174-2022-2-22.
Growcoot, Matt. “Meta’s AI Image Generator Accused of Making ‘Woke’ Pictures Like Google.” PetaPixel, March 4, 2024. https://petapixel.com/2024/03/04/metas-ai-image-generator-accused-of-making-woke-pictures-like-google/.
Guess, Andrew, Jonathan Nagler, and Joshua Tucker. “Less than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook.” *Science Advances* 5, no. 1 (January 9, 2019): 4586. https://doi.org/10.1126/sciadv.aau4586.
Habermas, Jürgen. “Reflections and Hypotheses on a Further Structural Transformation of the Political Public Sphere.” *Theory, Culture & Society* 39, no. 4 (July 1, 2022): 145–71. https://doi.org/10.1177/02632764221112341.
———. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Translated by Thomas Burger. Cambridge, MA: MIT Press, 1989.
Habgood-Coote, Joshua. “Deepfakes and the Epistemic Apocalypse.” *Synthese* 201, no. 3 (March 9, 2023): 103. https://doi.org/10.1007/s11229-023-04097-3.
Haggerty, Kevin D., and Richard V. Ericson. “The Surveillant Assemblage.” *The British Journal of Sociology* 51, no. 4 (2000): 605–22. https://doi.org/10.1080/00071310020015280.
Halevy, Alon, Peter Norvig, and Fernando Pereira. “The Unreasonable Effectiveness of Data.” *IEEE Intelligent Systems* 24, no. 2 (March 2009): 8–12. https://doi.org/10.1109/MIS.2009.36.
Hall, Stuart. *Representation: Cultural Representations and Signifying Practices*. 1st ed. Culture, Media and Identities Series. Sage Publications & Open University, 1997.
Hall, Stuart, Chas Critcher, Tony Jefferson, John Clarke, and Brian Roberts. *Policing the Crisis: Mugging, the State, and Law and Order*. The Macmillan Press, 1978.
Hameleers, Michael. “Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting.” *International Journal of Public Opinion Research* 36, no. 1 (March 1, 2024): edae004. https://doi.org/10.1093/ijpor/edae004.
Harcourt, Bernard E. *The Counterrevolution: How Our Government Went to War Against Its Own Citizens*. Basic Books, 2018.
Hardy, Jonathan. *Critical Political Economy of the Media: An Introduction*. 1st ed. Communication and Society. Routledge, 2014.
Harris, Keith Raymond and Philosophy Documentation Center. “Epistemic Domination.” *Thought: A Journal of Philosophy*, 2023. https://doi.org/10.5840/tht202341317.
Harsin, Jayson. “Regimes of Posttruth, Postpolitics, and Attention Economies.” *Communication, Culture & Critique* 8, no. 2 (2015): 327–33. https://doi.org/10.1111/cccr.12097.
Hartmann, Jochen, Jasper Schwenzow, and Maximilian Witte. “The Political Ideology of Conversational AI: Converging Evidence on ChatGPT’s pro-Environmental, Left-Libertarian Orientation.” SSRN Scholarly Paper. Rochester, NY, January 1, 2023. https://doi.org/10.2139/ssrn.4316084.
Harvey, Adam. “Circular Diffusion.” *Adam Harvey Studio* (blog), August 21, 2022. https://adam.harvey.studio/circular-diffusion/.
Haugeland, John. *Artificial Intelligence: The Very Idea*. 1st ed. MIT Press, 1985.
Hayles, N. Katherine. “Subversion of the Human Aura: A Crisis in Representation.” *American Literature* 95, no. 2 (June 1, 2023): 255–79. https://doi.org/10.1215/00029831-10575063.
Hayward, Tim. “The Problem of Disinformation: A Critical Approach.” *Social Epistemology* 0, no. 0 (2024): 1–23.

https://doi.org/10.1080/02691728.2024.2346127.
Heaven, Will. “Large Language Models Can Do Jaw-Dropping Things. But Nobody Knows Exactly Why.” MIT Technology Review, March 4, 2024. https://www.technologyreview.com/2024/03/04/1089403/large-language-models-amazing-but-nobody-knows-why/.
Heidegger, Martin. “The Age of The World Picture” in *The Question Concerning Technology and Other Essays* (New York: Harper Torchbooks, 1977).
Hense, Svenja, and Armin Schäfer. “Unequal Representation and the Right-Wing Populist Vote.” In *Contested Representation: Challenges, Shortcomings and Reforms*, edited by Claudia Landwehr, Thomas Saalfeld, and Armin Schäfer. Cambridge University Press, 2022.
Herman, Edward S., and Noam Chomsky. *Manufacturing Consent: The Political Economy of the Mass Media*. Pantheon Books, 1988.
Hern, Alex. “Facebook: We Were Too Slow to Recognise Our ‘corrosive’ Effect on Democracy.” *The Guardian*, January 22, 2018, sec. Technology. https://www.theguardian.com/technology/2018/jan/22/facebook-too-slow-social-media-fake-news-hiring.
Hernández-Orallo, José. *The Measure of All Minds: Evaluating Natural and Artificial Intelligence*. Cambridge: Cambridge University Press, 2017. https://doi.org/10.1017/9781316594179.
Hillebrandt, Maarten. “The Communicative Model of Disinformation: A Literature Note.” SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, March 5, 2021. https://doi.org/10.2139/ssrn.3798729.
Hinton, Geoffrey E. “Connectionist Learning Procedures.” *Artificial Intelligence* 40, no. 1 (September 1, 1989): 185–234. https://doi.org/10.1016/0004-3702(89)90049-0.
Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. “A Fast Learning Algorithm for Deep Belief Nets.” *Neural Computation* 18, no. 7 (July 2006): 1527–54. https://doi.org/10.1162/neco.2006.18.7.1527.
Hirsh, Jesse. “128: The End of the Public?” Substack newsletter. *Metaviews: Future of Authority* (blog), March 12, 2025. https://metaviews.substack.com/p/the-end-of-the-public?utm_medium=web.
Ho, Jonathan, Ajay Jain, and Pieter Abbeel. “Denoising Diffusion Probabilistic Models.” arXiv, December 16, 2020. https://doi.org/10.48550/arXiv.2006.11239.
Hofmann, Valentin, Pratyusha Ria Kalluri, Dan Jurafsky, and Sharese King. “Dialect Prejudice Predicts AI Decisions about People’s Character, Employability, and Criminality.” arXiv, March 1, 2024. https://doi.org/10.48550/arXiv.2403.00742.
Hooker, Lucy, and Chris Vallance. “Google Ending AI Arms Ban Incredibly Concerning, Campaigners Say.” BBC News, February 5, 2025. https://www.bbc.com/news/articles/cy081nqx2zjo.
Hou, Rui, and Diana Fu. “Sorting Citizens: Governing via China’s Social Credit System.” *Governance* 37, no. 1 (2024): 59–78. https://doi.org/10.1111/gove.12751.
House, The White. “FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI.” The White House, July 21, 2023. https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.
Howard, Jeremy. “AI Safety and the Age of Dislightenment.” fast.ai, July 10, 2023. https://www.fast.ai/posts/2023-11-07-dislightenment.html#abstract.
Huang, Saffron, and Divya Siddarth. “Generative AI and the Digital Commons.” arXiv, March 20, 2023. https://doi.org/10.48550/arXiv.2303.11074.
Hupkes, Dieuwke, Mario Giulianelli, Verna Dankers, Mikel Artetxe, Yanai Elazar, Tiago Pimentel, Christos Christodoulopoulos, et al. “A Taxonomy and Review of Generalization Research in NLP.” *Nature Machine Intelligence* 5, no. 10 (October 2023): 1161–74. https://doi.org/10.1038/s42256-023-00729-y.
Hyska, Megan. “The Politics of Past and Future: Synthetic Media, Showing, and Telling.” *Philosophical Studies* 182, no. 1 (January 2025): 137–58. https://doi.org/10.1007/s11098-023-02062-x.
Inc, Gallup. “Confidence in Institutions.” Gallup.com, 2024. https://news.gallup.com/poll/1597/Confidence-Institutions.aspx.
Ingram, Mathew. “The Supreme Court Hears Arguments about Government ‘Jawboning.’” Columbia Journalism Review. Accessed December 14, 2024. https://www.cjr.org/the_media_today/supreme_court_louisiana_missouri_social_media.php.
Innis, Harold Adams. *Empire and Communications*. Clarendon Press, 1950.
Jack Murphy Live. “J.D. Vance - JML #070,” September 17, 2021. Youtube. https://www.youtube.com/watch?v=PMq1ZEcyztY.
Jacobsen, Benjamin N. “Regimes of Recognition on Algorithmic Media.” *New Media & Society* 25, no. 12

(December 1, 2023): 3641–56. https://doi.org/10.1177/14614448211053555.
Jager, Anton. “From Post-Politics to Hyper-Politics.” Jacobin, February 14, 2022. https://jacobin.com/2022/02/from-post-politics-to-hyper-politics.
Jain, Niharika, Lydia Manikonda, Alberto Olmo Hernandez, Sailik Sengupta, and Subbarao Kambhampati. “Imagining an Engineer: On GAN-Based Data Augmentation Perpetuating Biases.” arXiv, November 9, 2018. https://doi.org/10.48550/arXiv.1811.03751.
Jakesch, Maurice, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. “Co-Writing with Opinionated Language Models Affects Users’ Views.” In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–15. CHI ’23. New York, NY, USA: Association for Computing Machinery, 2023. https://doi.org/10.1145/3544548.3581196.
Jasper, Susan. “How We’re Approaching the 2024 U.S. Elections.” *The Keyword* (blog), December 19, 2023. https://blog.google/outreach-initiatives/civics/how-were-approaching-the-2024-us-elections/.
Ji, Jiaming, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan, et al. “AI Alignment: A Comprehensive Survey.” arXiv, February 26, 2024. http://arxiv.org/abs/2310.19852.
Jiancai, Chen. “Theoretical Communication Needs to Have a Strong Trunk and a Flourishing Crown,” *网络传播 (New Media)*, February 2016. https://www.cac.gov.cn/files/pdf/zazhiwlcb/zazhiwlcb201602.pdf
Jinping, Xi. “Secure a Decisive Victory in Building a Moderately Prosperous Society in All Respects and Strive for the Great Success of Socialism with Chinese Characteristics for a New Era,” transcript of speech delivered at the 19th National Congress of the Communist Party of China, Beijing, October 18, 2017, http://www.xinhuanet.com/english/download/Xi_Jinping's_report_at_19th_CPC_National_Congress.pdf
Jones, Phil. “Make It New.” *Verso* (blog), June 20, 2023. https://www.versobooks.com/en-ca/blogs/news/make-it-new.
Joque, Justin. *Revolutionary Mathematics: Artificial Intelligence, Statistics and the Logic of Capitalism*. Verso Books, 2022.
Jungherr, Andreas, and Adrian Rauchfleisch. “Negative Downstream Effects of Alarmist Disinformation Discourse: Evidence from the United States.” *Political Behavior*, January 12, 2024. https://doi.org/10.1007/s11109-024-09911-3.
Jungherr, Andreas, and Ralph Schroeder. “Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy.” *Social Media + Society* 7, no. 1 (January 1, 2021): 2056305121988928. https://doi.org/10.1177/2056305121988928.
Karabulut, Dogus, Cagri Ozcinar, and Gholamreza Anbarjafari. “Automatic Content Moderation on Social Media.” *Multimedia Tools and Applications* 82, no. 3 (January 1, 2023): 4439–63. https://doi.org/10.1007/s11042-022-11968-3.
Karpf, David. *The MoveOn Effect: The Unexpected Transformation of American Political Advocacy*. Oxford University Press, 2012.
Karras, Tero, Samuli Laine, and Timo Aila. “A Style-Based Generator Architecture for Generative Adversarial Networks.” arXiv, March 29, 2019. https://doi.org/10.48550/arXiv.1812.04948.
Kasy, Maximilian. “The Political Economy of AI: Towards Democratic Control of the Means of Prediction.” Discussion Paper Series. IZA Institute of Labour Economics, April 2024.
Katz, Yarden. *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. New York: Columbia University Press, 2020.
Kavada, Anastasia. “Creating the Collective: Social Media, the Occupy Movement and Its Constitution as a Collective Actor.” *Information, Communication & Society* 18, no. 8 (August 3, 2015): 872–86. https://doi.org/10.1080/1369118X.2015.1043318.
Kavanagh, Jennifer, and Michael D. Rich. “Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life.” RAND Corporation, January 15, 2018. https://www.rand.org/pubs/research_reports/RR2314.html.
Kay, Jackie, Atoosa Kasirzadeh, and Shakir Mohamed. “Epistemic Injustice in Generative AI.” *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 7 (October 16, 2024): 684–97. https://doi.org/10.1609/aies.v7i1.31671.
Kelleher, John D. *Deep Learning*. The MIT Press Essential Knowledge. MIT Press, 2019.
Khan, Mehtab, and Alex Hanna. “The Subjects and Stages of AI Dataset Development: A Framework for Dataset Accountability.” SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, September 13, 2022. https://doi.org/10.2139/ssrn.4217148.
Kilcullen, David. *Counterinsurgency*. Oxford University Press, USA, 2010.
Kingma, Diederik P., and Max Welling. “Auto-Encoding Variational Bayes.” arXiv, December 10, 2022.

https://doi.org/10.48550/arXiv.1312.6114.
Kirchner, Jan, Lama Ahmad, Scott Aaronson, and Jan Leike. “New AI Classifier for Indicating AI-Written Text.” Accessed April 24, 2024. https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text.
Kirkegaard, Jacob Funk. “China’s Population Decline Is Getting Close to Irreversible.” Peterson Institute for International Economics, January 18, 2024. https://www.piie.com/research/piie-charts/2024/chinas-population-decline-getting-close-irreversible.
Kiss, Nandor. “The Twenty-Six Words That Created the Internet… and Then Maybe, Kind Of, Destroyed Society: Understanding and Reforming Section 230 of the Communications Decency Act.” *Michigan Technology Law Review* 31, no. 1 (January 1, 2024): 131–65.
Kleinman, Zoe, and Liv McMahon. “UK and US Refuse to Sign International AI Declaration.” BBC News, February 11, 2025. https://www.bbc.com/news/articles/c8edn0n58gwo.
Klincewicz, Michal, Mark Alfano, and Amir Fard. “Slopaganda: The Interaction Between Propaganda and Generative Ai.” *Filosofiska Notiser* 12, no. 1 (2025): 135–62.
Knorr Cetina, Karin. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, UNITED STATES: Harvard University Press, 1999.
Krawetz, Neal. “C2PA’s Butterfly Effect,’ Hacker Factor, November 16, 2023, accessed February 26, 2024, https://www.hackerfactor.com/blog/index.php?/archives/1010-C2PAs-Butterfly-Effect.html
———. “C2PA’s Worst Case Scenario,’ Hacker Factor, December 18, 2023, accessed February 26, 2024, https://www.hackerfactor.com/blog/index.php?/archives/1013-C2PAs-Worst-Case-Scenario.html
Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet Classification with Deep Convolutional Neural Networks.” In *Advances in Neural Information Processing Systems*, Vol. 25. Curran Associates, Inc., 2012. https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.
Kruppa, Miles. “Jeff Bezos Bets on a Google Challenger Using AI to Try to Upend Internet Search.” WSJ, January 4, 2024. https://www.wsj.com/tech/ai/jeff-bezos-bets-on-a-google-challenger-using-ai-to-try-to-upend-internet-search-0859bda6.
Krzanowski, Roman, and Pawel Polak. “The Internet as an Epistemic Agent (EA).” *Információs Társadalom* 22, no. 2 (August 31, 2022): 39. https://doi.org/10.22503/inftars.XXII.2022.2.3.
Kuchlous, Sahil, Marvin Li, and Jeffrey G. Wang. “Bias Begets Bias: The Impact of Biased Embeddings on Diffusion Models.” arXiv, September 15, 2024. https://doi.org/10.48550/arXiv.2409.09569.
Kumar, Deepak, Yousef AbuHashem, and Zakir Durumeric. “Watch Your Language: Investigating Content Moderation with Large Language Models.” arXiv, January 17, 2024. https://doi.org/10.48550/arXiv.2309.14517.
Kuo, Rachel, and Alice Marwick. “Critical Disinformation Studies: History, Power, and Politics.” *Harvard Kennedy School Misinformation Review*, August 12, 2021. https://doi.org/10.37016/mr-2020-76.
Labbe, Mark. “The Deepfake 2020 Election Threat Is Real, but Containable.” TechTarget, September 25, 2020. https://www.techtarget.com/searchenterpriseai/feature/The-deepfake-2020-election-threat-is-real-but-containable.
Laney, Doug. “3D Data Management: Controlling Data Volume, Velocity and Variety.” Research Note. Application Delivery Strategies. META Group Inc., February 6, 2001. https://diegonogare.net/wp-content/uploads/2020/08/3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.
Lasswell, Harold D. “The Theory of Political Propaganda.” *The American Political Science Review* 21, no. 3 (1927): 627–31. https://doi.org/10.2307/1945515.
Latour, Bruno. *Pandora’s Hope: Essays on the Reality of Science Studies*. 2. print. Cambridge, Mass.: Harvard University Press, 2000.
———. *Reassembling the Social. An Introduction to Actor-Network-Theory*. Oxford University Press, 2005.
Leary, Kyree. “An AI That Makes Fake Videos May Facilitate the End of Reality as We Know It.” Futurism, December 8, 2017. https://futurism.com/ai-makes-fake-videos-facilitate-end-reality-know-it.
LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. “Deep Learning.” *Nature* 521, no. 7553 (May 2015): 436–44. https://doi.org/10.1038/nature14539.
Lee, Nicol, and Obioha Chijioke. “Why States and Localities Are Acting on AI.” Brookings. Accessed March 15, 2024. https://www.brookings.edu/articles/why-states-and-localities-are-acting-on-ai/.
Leese, Matthias. “‘Seeing Futures’: Politics of Visuality and Affect.” In *Algorithmic Life*. Routledge, 2016.
Lemley, Mark A., and Bryan Casey. “Fair Learning.” SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, January 30, 2020. https://doi.org/10.2139/ssrn.3528447.
Lenoir, Théophile, and Chris Anderson. “Introduction Essay: What Comes After Disinformation Studies.” *Center*

for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill, January 23, 2023. https://citap.pubpub.org/pub/oijfl3sv/release/1.

Lessig, Lawrence. *Code: And Other Laws of Cyberspace, Version 2.0*. Basic Books, 2006.

Li, Cathy, and Agustina Callegari. "How AI Can Also Be Used to Combat Online Disinformation." World Economic Forum, June 14, 2024. https://www.weforum.org/stories/2024/06/ai-combat-online-misinformation-disinformation/.

Li, Yuanzhi, Sébastien Bubeck, Ronen Eldan, Allie Del Giorno, Suriya Gunasekar, and Yin Tat Lee. "Textbooks Are All You Need II: Phi-1.5 Technical Report." arXiv, September 11, 2023. https://doi.org/10.48550/arXiv.2309.05463.

Liedke, Christopher St Aubin and Jacob. "-Most Americans Favor Restrictions on False Information, Violent Content Online." *Pew Research Center* (blog), July 20, 2023. https://www.pewresearch.org/short-reads/2023/07/20/most-americans-favor-restrictions-on-false-information-violent-content-online/.

Lim, Gabrielle. "SECURITIZE/ COUNTER-SECURITIZE." New York, NY: Data & Society, March 25, 2020.

Lippmann, Walter. *Public Opinion*. Routledge, 1997.

Liu, Yifei, Yuang Panwang, and Chao Gu. "'Turning Right'? An Experimental Study on the Political Value Shift in Large Language Models." *Humanities and Social Sciences Communications* 12, no. 1 (February 10, 2025): 1–10. https://doi.org/10.1057/s41599-025-04465-z.

Liu, Zongyuan Zoe. "China's Real Economic Crisis." *Foreign Affairs*, August 6, 2024. https://www.foreignaffairs.com/china/chinas-real-economic-crisis-zongyuan-liu.

Llansó, Emma J. "No Amount of 'AI' in Content Moderation Will Solve Filtering's Prior-Restraint Problem." *Big Data & Society* 7, no. 1 (January 1, 2020): 6. https://doi.org/10.1177/2053951720920686.

Lucas D. Introna, Helen Nissenbaum. "Shaping the Web: Why the Politics of Search Engines Matters." *The Information Society* 16, no. 3 (July 1, 2000): 169–85. https://doi.org/10.1080/01972240050133634.

Luccioni, Alexandra Sasha, Christopher Akiki, Margaret Mitchell, and Yacine Jernite. "Stable Bias: Analyzing Societal Representations in Diffusion Models." arXiv, November 9, 2023. https://doi.org/10.48550/arXiv.2303.11408.

Luciano, Floridi. "Hypersuasion – On AI's Persuasive Power and How to Deal with It." *Philosophy & Technology* 37, no. 2 (May 17, 2024): 64. https://doi.org/10.1007/s13347-024-00756-6.

Lunden, Ingrid. "UK Drops 'safety' from Its AI Body, Now Called AI Security Institute, Inks MOU with Anthropic." *TechCrunch* (blog), February 14, 2025. https://techcrunch.com/2025/02/13/uk-drops-safety-from-its-ai-body-now-called-ai-security-institute-inks-mou-with-anthropic/.

Mac History, *Apple Knowledge Navigator Video (1987)*, 2012. https://www.youtube.com/watch?v=umJsITGzXd0.

Mackenzie, Adrian. *Machine Learners: Archaeology of a Data Practice*. MIT Press. The MIT Press, 2017.

MacKenzie, Adrian, and Anna Munster. "Platform Seeing: Image Ensembles and Their Invisualities." *Theory, Culture & Society* 36, no. 5 (September 1, 2019): 3–22. https://doi.org/10.1177/0263276419847508.

Maluleke, Vongani H., Neerja Thakkar, Tim Brooks, Ethan Weber, Trevor Darrell, Alexei A. Efros, Angjoo Kanazawa, and Devin Guillory. "Studying Bias in GANs through the Lens of Race." arXiv, September 14, 2022. http://arxiv.org/abs/2209.02836.

Manfredi-Sánchez, Juan Luis, and Pablo Sebastian Morales. "Generative AI and the Future for China's Diplomacy." *Place Branding and Public Diplomacy*, March 11, 2024. https://doi.org/10.1057/s41254-024-00328-7.

Marcus, Gary. "What If Generative AI Turned out to Be a Dud?" Substack newsletter. *Marcus on AI* (blog), August 12, 2023. https://garymarcus.substack.com/p/what-if-generative-ai-turned-out.

Markov, Todor, Chong Zhang, Sandhini Agarwal, Tyna Eloundou, Teddy Lee, Steven Adler, Angela Jiang, and Lilian Weng. "A Holistic Approach to Undesired Content Detection in the Real World." arXiv, February 14, 2023. http://arxiv.org/abs/2208.03274.

Marx, Karl. "Estranged Labor," in *Economic and Philosophic Manuscripts of 1844*, trans. Martin Milligan (Progress Publishers, 1959).

Matt Rickard. "A List of Leaked System Prompts." Blog, May 17, 2023. https://matt-rickard.com/a-list-of-leaked-system-prompts.

McKendrick, Kathleen. "Artificial Intelligence: Prediction and Counterterrorism." Research Report. London: The Royal Institute of International Affairs, August 2019. https://www.chathamhouse.org/sites/default/files/2019-08-07-AICounterterrorism.pdf.

McLoughlin, Killian L., and William J. Brady. "Human-Algorithm Interactions Help Explain the Spread of Misinformation." *Current Opinion in Psychology* 56 (April 1, 2024): 101770. https://doi.org/10.1016/j.copsyc.2023.101770.

McPhedran, Robert, Michael Ratajczak, Max Mawby, Emily King, Yuchen Yang, and Natalie Gold. "Psychological

Inoculation Protects against the Social Media Infodemic." *Scientific Reports* 13, no. 1 (April 8, 2023): 5780. https://doi.org/10.1038/s41598-023-32962-1.
"Meet Argus: Accrete's Advanced AI Threat Detection Agent." Accessed May 29, 2024. https://www.accrete.ai/argus.
Mejia, Robert, Kay Beckermann, and Curtis Sullivan. "White Lies: A Racial History of the (Post)Truth." *Communication and Critical/Cultural Studies* 15, no. 2 (April 3, 2018): 109–26. https://doi.org/10.1080/14791420.2018.1456668.
Mejias, Ulises Ali, and Nick Couldry. *Data Grab: The New Colonialism of Big Tech and How to Fight Back*. Chicago, IL: The University of Chicago Press, 2024.
"Membership - C2PA." Accessed January 15, 2025. https://c2pa.org/membership/.
Memon, Shahan Ali, and Jevin D. West. "Search Engines Post-ChatGPT: How Generative Artificial Intelligence Could Make Search Less Reliable." arXiv, February 18, 2024. http://arxiv.org/abs/2402.11707.
Meng, Amanda, and Carl DiSalvo. "Grassroots Resource Mobilization through Counter-Data Action." *Big Data & Society* 5, no. 2 (July 1, 2018): 2053951718796862. https://doi.org/10.1177/2053951718796862.
Menon, Sachit, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. "PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models." arXiv, July 20, 2020. https://doi.org/10.48550/arXiv.2003.03808.
Merrill, Jeremy, and Will Oremus. "Five Points for Anger, One for a 'like': How Facebook's Formula Fostered Rage and Misinformation." Washington Post, October 26, 2021. https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/.
Mesquita, Ethan, Brandice Canes-Wrone, Andrew Hall, Kristian Lum, Gregory Martin, and Yamil Velez. "Preparing for Generative AI in the 2024 Election: Recommendations and Best Practices Based on Academic Research." White paper. Chicago: University of Chicago Harris School of Public Policy, August 2023.
Meta. "Hard Questions: What's Facebook's Strategy for Stopping False News?," May 23, 2018. https://about.fb.com/news/2018/05/hard-questions-false-news/.
———. "Partnering with Third-Party Fact-Checkers." *Meta Journalism Project* (blog), March 23, 2020. https://www.facebook.com/journalismproject/programs/third-party-fact-checking/selecting-partners.
———. "Safeguards for Elections," *Meta*, Accessed February 15, 2024. https://about.meta.com/actions/preparing-for-elections-with-meta/.
———. "What's New Across Our AI Experiences," Meta, December 6, 2023. https://about.fb.com/news/2023/12/meta-ai-updates/.
MeVer. "MeVer," *MeVer*. Accessed April 20, 2024. https://mever.gr/services/
Meyer, Roland. "It's a Flat World. The Synthetic Realities of Sora." Application/pdf. *Rrrreflect. Journal of Integrated Design Research; Special Issue 1*, 2024, 151 KB, 4 pages. https://doi.org/10.57684/COS-1267.
Miceli, Milagros, and Julian Posada. "The Data-Production Dispositif." arXiv, May 24, 2022. https://doi.org/10.48550/arXiv.2205.11963.
Ministry of Foreign Affairs. *Global AI Governance Action Plan*. July 26, 2025. https://www.mfa.gov.cn/zyxw/202507/t20250726_11677803.shtml Ministry of Foreign Affairs of Japan. "G7 Leaders' Statement on the Hiroshima AI Process." Accessed October 28, 2024. https://www.mofa.go.jp/ecm/ec/page5e_000076.html.
Mintz, Jocelyn. "'Garbage in, Garbage out': AI Fails to Debunk Disinformation, Study Finds." Voice of America, October 21, 2024. https://www.voanews.com/a/garbage-in-garbage-out-ai-fails-to-debunk-disinformation-study-finds/7830414.html.
Mira, José Mira. "Symbols versus Connections: 50 Years of Artificial Intelligence." *Neurocomputing*, Neural Networks: Algorithms and Applications, 71, no. 4 (January 1, 2008): 671–80. https://doi.org/10.1016/j.neucom.2007.06.009.
Mirzoeff, Nicholas. "On Visuality." *Journal of Visual Culture* 5 (April 1, 2006): 53–79. https://doi.org/10.1177/1470412906062285.
———. *The Right to Look: A Counterhistory of Visuality*. Durham: Duke University Press, 2011.
Monett, Dagmar, and Colin W. P. Lewis. "Getting Clarity by Defining Artificial Intelligence—A Survey." edited by Vincent C. Müller, 44:212–14. Studies in Applied Philosophy, Epistemology and Rational Ethics. Berlin: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-96448-5_21.
Montasari, Reza. "The Dual Role of Artificial Intelligence in Online Disinformation: A Critical Analysis." In *Cyberspace, Cyberterrorism and the International Security in the Fourth Industrial Revolution: Threats, Assessment and Responses*, edited by Reza Montasari, 229–40. Cham: Springer International Publishing,

2024. https://doi.org/10.1007/978-3-031-50454-9_11.
Mosco, Vincent. *The Political Economy of Communication: Rethinking and Renewal*. SAGE Publications, 1996.
Mosseri, Adam. "Bringing People Closer Together." *Meta* (blog), January 12, 2018. https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/.
Mukjerjee, Supantha. "Explainer: Will the EU Delay Enforcing Its AI Act? | Reuters." Reuters, July 3, 2025. https://www.reuters.com/business/media-telecom/will-eu-delay-enforcing-its-ai-act-2025-07-03/.
Mulligan, Scott. "AI Trained on AI Garbage Spits out AI Garbage." MIT Technology Review, July 24, 2024. https://www.technologyreview.com/2024/07/24/1095263/ai-that-feeds-on-a-diet-of-ai-garbage-ends-up-spitting-out-nonsense/.
Muñoz, José Esteban. *Cruising Utopia: The Then and There of Queer Futurity*. NYU Press, 2009. https://www.jstor.org/stable/j.ctt9qg4nr.
Murdock, Graham, and Peter Golding. "For a Political Economy of Mass Communications." *Socialist Register* 10 (March 18, 1973). https://socialistregister.com/index.php/srv/article/view/5355.
Mutsvairo, Bruce, Luca Bruls, Mirjam de Bruijn, and Kristin Skare Orgeret. "'Factforward:' Foretelling the Future of Africa's Information Disorder." *Center for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill*, January 23, 2023. https://citap.pubpub.org/pub/lopohgbv/release/1.
Nadeem, Mohammad, Shahab Saquib Sohail, Erik Cambria, Björn W. Schuller, and Amir Hussain. "Gender Bias in Text-to-Video Generation Models: A Case Study of Sora." arXiv, January 10, 2025. https://doi.org/10.48550/arXiv.2501.01987.
National Security Council, "National Strategy for Countering Domestic Terrorism," June 2021, 8. https://www.whitehouse.gov/wp-content/uploads/2021/06/National-Strategy-for-Countering-DomesticTerrorism.pdf
"Nebula Social™ by Accrete.AI." Accrete. Accessed May 29, 2024. https://www.nebulasocial.ai/.
Negroponte, Nicholas. *Being Digital*. London: Hodder & Stoughton, 1995.
"New Technologies and the Christchurch Call: Challenges, Mitigations, and Opportunities." Christchurch Call Foundation, 2023. https://www.christchurchcall.com/assets/Summit-23/Leaders-Summit-2023-Meeting-paper-New-technologies-and-the-Christchurch-Call-Challenges-mitigations-and-opportunities.pdf.
Ngak, Chenda. "Occupy Wall Street Uses Social Media to Spread Nationwide - CBS News," October 13, 2011. https://www.cbsnews.com/news/occupy-wall-street-uses-social-media-to-spread-nationwide/.
Nicholas, Gabriel. "Shedding Light on Shadowbanning." Washington, D.C: Center for Democracy & Technology, May 19, 2022. https://osf.io/xcz2t.
Nick Couldry, Ulises A. Mejias. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. 1st ed. Culture and Economic Life. Redwood City: Stanford University Press, 2019.
Ninghui, Sun. "The Development of Artificial Intelligence and Intelligent Computing," *Lecture Notes for the Tenth Special Lecture of the 14th National People's Congress Standing Committee*, April 2024, http://www.npc.gov.cn/npc/c2/c30834/202404/t20240430_436915.html
Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press, 2018.
NPR Staff. "Occupy Wall Street Inspires Worldwide Protests." *NPR*, October 15, 2011, sec. World. https://www.npr.org/2011/10/15/141382468/occupy-wall-street-inspires-worldwide-protests.
NVIDIA. "NVIDIA Keynote at SIGGRAPH 2023," Youtube. Aug. 8 2023. 20:00-20:42. https://www.youtube.com/watc,h?v=Z2VBKerS63A.
NVIDIA Research Projects. "Flickr-Faces-HQ Dataset (FFHQ)." GitHub Repository. https://github.com/NVlabs/ffhq-dataset/.
Nyhan, Brendan. "Facts and Myths about Misperceptions." *Journal of Economic Perspectives* 34, no. 3 (August 2020): 220–36. https://doi.org/10.1257/jep.34.3.220.
Offert, Fabian, and Thao Phan. "A Sign That Spells: DALL-E 2, Invisual Images and The Racial Politics of Feature Space." arXiv, October 26, 2022. https://doi.org/10.48550/arXiv.2211.06323.
Olay, Matthew. "Senior Officials Outline President's Proposed FY26 Defense Budget." *DoD News* (blog), June 26, 2025. https://www.defense.gov/News/News-Stories/Article/Article/4227847/senior-officials-outline-presidents-proposed-fy26-defense-budget/.
O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016.
OpenAI. "Aligning Language Models to Follow Instructions," OpenAI, January 27, 2022, https://openai.com/research/instruction-following.
———. "DallE-3 System Card," OpenAI, October 3, 2023,

https://cdn.openai.com/papers/DALL_E_3_System_Card.pdf
———. “GPT-4o System Card,” August 8, 2024. https://openai.com/index/gpt-4o-system-card/.
———. “How OpenAI Is Approaching 2024 Worldwide Elections.” Accessed November 26, 2024. https://openai.com/index/how-openai-is-approaching-2024-worldwide-elections/.
———. “Introducing Superalignment,” OpenAI, July 5, 2023, https://openai.com/blog/introducingsuperalignment
———. “Moderations” OpenAI Platform, accessed April 8, 2024, https://platform.openai.com.docs/apireference/moderation
———. “OpenAI Appoints Retired U.S. Army General,” June 13, 2024. https://openai.com/index/openai-appoints-retired-us-army-general/.

———. “OpenAI Platform.” Accessed April 8, 2024. https://platform.openai.com.
Orben, Amy. “The Sisyphean Cycle of Technology Panics.” *Perspectives on Psychological Science* 15, no. 5 (September 1, 2020): 1143–57. https://doi.org/10.1177/1745691620919372.
O’Reilly, Gillian Hadfield, Mariano-Florentino (Tino) Cuéllar, Tim. “It’s Time to Create a National Registry for Large AI Models.” Carnegie Endowment for International Peace. Accessed March 21, 2024. https://carnegieendowment.org/2023/07/12/it-s-time-to-create-national-registry-for-large-ai-models-pub-90180.
Osmundsen, Mathias, Alexander Bor, Peter Bjerregaard Vahlstrup, Anja Bechmann, and Michael Bang Petersen. “Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter.” *American Political Science Review* 115, no. 3 (August 2021): 999–1015. https://doi.org/10.1017/S0003055421000290.
Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, et al. “Training Language Models to Follow Instructions with Human Feedback.” arXiv, March 4, 2022. http://arxiv.org/abs/2203.02155.
Oxford Languages. “Oxford Word of the Year 2016.” Accessed December 6, 2024. https://languages.oup.com/word-of-the-year/2016/.
Paglen, Trevor. “Invisible Images (Your Pictures Are Looking at You).” *The New Inquiry* (blog), December 8, 2016. https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/.
Paris, Britt. “Seeing Through the Fog of War: Assessing Epistemic Burden Around Cheapfakes and Deepfakes of Geopolitical Crisis.” In *Re-Thinking Mediations of Post-Truth Politics and Trust*. Routledge, 2023.
Paris, Britt, and Joan Donovan. “Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence.” *Data & Society*, September 2019.
Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. 1st ptg. Penguin Press HC, 2011.
Pasquale, Frank. “Introduction: The Need to Know.” In *The Black Box Society*, 1–18. The Secret Algorithms That Control Money and Information. Harvard University Press, 2015.
Pasquinelli, Matteo. *The Eye of the Master: A Social History of Artificial Intelligence*. London: Verso Books, 2023.
———. “Three Thousand Years of Algorithmic Rituals: The Emergence of AI from the Computation of Space.” *E-Flux*, no. 101 (June 2019). https://www.e-flux.com/journal/101/273221/three-thousand-years-of-algorithmic-rituals-the-emergence-of-ai-from-the-computation-of-space/.
Pasquinelli, Matteo, and Vladan Joler. “The Nooscope Manifested: AI as Instrument of Knowledge Extractivism.” *AI & SOCIETY* 36, no. 4 (December 2021): 1263–80. https://doi.org/10.1007/s00146-020-01097-6.
Patel, Faiza. “U.S. AI-Driven ‘Catch and Revoke’ Initiative Threatens First Amendment Rights | Brennan Center for Justice.” Brennan Center for Justice, March 18, 2025. https://www.brennancenter.org/our-work/analysis-opinion/us-ai-driven-catch-and-revoke-initiative-threatens-first-amendment-rights.
Paullada, Amandalynne, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. “Data and Its (Dis)Contents: A Survey of Dataset Development and Use in Machine Learning Research.” arXiv, December 9, 2020. https://doi.org/10.48550/arXiv.2012.05345.
Pennycook, Gordon, Adam Bear, Evan T. Collins, and David G. Rand. “The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings.” *Management Science* 66, no. 11 (November 2020): 4944–57. https://doi.org/10.1287/mnsc.2019.3478.
Pennycook, Gordon, and David G. Rand. “Accuracy Prompts Are a Replicable and Generalizable Approach for Reducing the Spread of Misinformation.” *Nature Communications* 13, no. 1 (April 28, 2022): 2333. https://doi.org/10.1038/s41467-022-30073-5.
Pentland, Alex, J. Penn, Christina Colclough, and Thomas Hardjono. “Data Cooperatives: Digital Empowerment of Citizens and Workers.” Whitepaper. MIT Connection Science. Massachusetts Institute of Technology,

January 2, 2019. https://ide.mit.edu/sites/default/files/publications/Data-Cooperatives-final.pdf.
Pereira, David. “Bias in AI: Much more than a Data problem.” Towards Data Science (blog), July 6, 2020. https://towardsdatascience.com/bias-in-ai-much-more-than-a-data-problem-de6ef950c848/.
Platformer. “Google’s AI Search Setback,” May 29, 2024. https://www.platformer.news/google-ai-overviews-eat-rocks-glue-pizza/.
Plato, “Allegory of the Cave” in *The Republic* (London: Penguin Books, 1987).
Politico. “‘Existential to Who?’ US VP Kamala Harris Urges Focus on near-Term AI Risks,” November 1, 2023. https://www.politico.eu/article/existential-to-who-us-vp-kamala-harris-urges-focus-on-near-term-ai-risks/.
Porpora, Douglas, and Seif Sekalala. “Truth, Communication, and Democracy.” *International Journal of Communication* 13, no. 0 (February 26, 2019): 18.
Prime Minister of Canada. “Joint Statement by Prime Minister Carney and Prime Minister Starmer,” June 15, 2025. https://www.pm.gc.ca/en/news/statements/2025/06/15/joint-statement-prime-minister-mark-carney-and-prime-minister-sir-keir-starmer.
Puutio, Alexander, and David A. Timis. “Deepfake Democracy: Here’s How Modern Elections Could Be Decided by Fake News.” World Economic Forum, October 5, 2020. https://www.weforum.org/stories/2020/10/deepfake-democracy-could-modern-elections-fall-prey-to-fiction/.
Quetelet, Lambert Adolphe Jacques. *A Treatise on Man and the Development of His Faculties*. W. & R. Chambers, 1842.
Raffel, Colin, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer.” arXiv, September 19, 2023. http://arxiv.org/abs/1910.10683.
Rai, Arun. “Explainable AI: From Black Box to Glass Box.” *Journal of the Academy of Marketing Science* 48, no. 1 (January 1, 2020): 137–41. https://doi.org/10.1007/s11747-019-00710-5.
Raley, Rita, and Jennifer Rhee. “Critical AI: A Field in Formation.” *American Literature* 95, no. 2 (June 1, 2023): 185–204. https://doi.org/10.1215/00029831-10575021.
Rancière, Jacques. *Hatred of Democracy*. Verso Books, 2014.
———. *The Politics of Aesthetics: The Distribution of the Sensible*. Continuum, 2004.
Reality Defender. “Technology — Reality Defender,” *Reality Defender*. Accessed April 20, 2024. https://realitydefender.com/technology
Rettenberger, Luca, Markus Reischl, and Mark Schutera. “Assessing Political Bias in Large Language Models.” arXiv, June 5, 2024. https://doi.org/10.48550/arXiv.2405.13041.
Reuters. “US Defense Department Awards Contracts to Google, Musk’s xAI.” July 14, 2025, sec. Autos & Transportation. https://www.reuters.com/business/autos-transportation/us-department-defense-awards-contracts-google-xai-2025-07-14/.
Richmond, Oliver P. “A Prelude to Revisionism? The Stalemated Peace Model and the Emergence of Multipolarity in International Order.” *Contemporary Security Policy*, April 3, 2025. https://www.tandfonline.com/doi/abs/10.1080/13523260.2024.2437928.
Rieder, Bernhard. “Scrutinizing an Algorithmic Technique: The Bayes Classifier as Interested Reading of Reality.” *Information, Communication & Society* 20, no. 1 (January 2, 2017): 100–117. https://doi.org/10.1080/1369118X.2016.1181195.
Rini, Regina. “Deepfakes and the Epistemic Backstop.” *Philosopher’s Imprint* 20, no. 24 (2020). http://hdl.handle.net/2027/spo.3521354.0020.024.
Riskin, Jessica. “The Defecating Duck, or, the Ambiguous Origins of Artificial Life.” *Critical Inquiry* 29, no. 4 (June 2003): 599–633. https://doi.org/10.1086/377722.
Roberge, Jonathan, and Michael Castelle, eds. *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*. Cham: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-56286-1.
Roberts, Tony, Judy Gitahi, Patrick Allam, Lawrence Oboh, and Oyewole Oladapo. “Mapping the Supply of Surveillance Technologies to Africa: Case Studies from Nigeria, Ghana, Morocco, Malawi, and Zambia.” Brighton: Institute of Development Studies, September 27, 2023. https://www.ids.ac.uk/publications/mapping-the-supply-of-surveillance-technologies-to-africa-case-studies-from-nigeria-ghana-morocco-malawi-and-zambia/.
Rosnay (Mélanie), Dulong de, and Stalder (Felix). “Digital Commons.” *Internet Policy Review* 9, no. 4 (December 17, 2020). https://doi.org/10.14763/2020.4.1530.
Routhier, Dominique, Lila Lee-Morrison, and Kathrin Maurer. “Automating Visuality: An Introduction.” *MAST The Journal of Media Art Study and Theory* 3, no. 1 (April 2022): 2–14.

Rouvroy, Antoinette, Thomas Berns, and Liz Carey-Libbrecht. “Algorithmic governmentality and prospects of emancipation: Disparateness as a precondition for individuation through relationships?” *Réseaux* 177, no. 1 (2013): 163–96.
Rozado, David. “Danger in the Machine: The Perils of Political and Demographic Biases Embedded in AI Systems.” Issue Brief. New York: The Manhattan Institute, March 2023.
Russell, Stuart. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking, 2019.
Russell, Stuart J., and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Third edition, Global edition. Prentice Hall Series in Artificial Intelligence. Boston Columbus Indianapolis: Pearson, 2016.
Sadowski, Jathan. “When Data Is Capital: Datafication, Accumulation, and Extraction.” *Big Data & Society* 6, no. 1 (January 1, 2019): 2053951718820549. https://doi.org/10.1177/2053951718820549.
Sag, Anshel. “AI Dominates Qualcomm Snapdragon Summit With New Snapdragon Products.” Forbes. Accessed February 13, 2024. https://www.forbes.com/sites/moorinsights/2023/10/25/ai-dominates-qualcomm-snapdragon-summit-with-new-snapdragon-products/.
Salakhutdinov, R., and Geoffrey E. Hinton. “Deep Boltzmann Machines.” In *International Conference on Artificial Intelligence and Statistics*, 2009. https://www.semanticscholar.org/paper/Deep-Boltzmann-Machines-Salakhutdinov-Hinton/ddc45fad8d15771d9f1f8579331458785b2cdd93.
Salazar, Grisel. “Rethinking Disinformation for the Global South: Towards a Particular Research Agenda.” In *State-Sponsored Disinformation Around the Globe*, edited by Martin Echeverria, Sara Santamaria, and Daniel Hallin. Routledge, 2024.
Salvi, Francesco, Manoel Horta Ribeiro, Riccardo Gallotti, and Robert West. “On the Conversational Persuasiveness of Large Language Models: A Randomized Controlled Trial.” arXiv, March 21, 2024. http://arxiv.org/abs/2403.14380.
Scheirer, Walter. *A History of Fake Things on the Internet*. 1st ed. Stanford University Press, 2023.
Scheuerman, Morgan Klaus, Alex Hanna, and Emily Denton. “Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development.” *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW2 (October 13, 2021): 1–37. https://doi.org/10.1145/3476058.
Schick, Nina. *Deepfakes: The Coming Infocalypse*. Grand Central Publishing, 2020.
Schramowski, Patrick, Manuel Brack, Björn Deiseroth, and Kristian Kersting. “Safe Latent Diffusion: Mitigating Inappropriate Degeneration in Diffusion Models.” arXiv, April 26, 2023. http://arxiv.org/abs/2211.05105.
Schreiner, Maximilian. “GPT-4 Architecture, Datasets, Costs and More Leaked.” THE DECODER, July 11, 2023. https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked/.
Sedlmeir, Johannes, Alexander Rieger, Tamara Roth, and Gilbert Fridgen. “Battling Disinformation with Cryptography.” *Nature Machine Intelligence* 5, no. 10 (October 2023): 1056–57. https://doi.org/10.1038/s42256-023-00733-2.
Seger, Elizabeth. “Should Epistemic Security Be a Priority GCR Cause Area.” *Intersections, Reinforcements, Cascades*, 2023, 18–37.
Sekula, Allan. “The Body and the Archive.” *October* 39 (1986): 3. https://doi.org/10.2307/778312.
Sensity AI, “The Best Deepfake Detection Solution | AI Image Detection,” *Sensity AI*, Accessed April 20, 2024. https://sensity.ai/deepfakes-detection/
Shahbaz, Adrian. “The Rise of Digital Authoritarianism.” Freedom House. Accessed October 25, 2024. https://freedomhouse.org/report/freedom-net/2018/rise-digital-authoritarianism.
Shahinfard, Farzad. “The Battle against AI-Driven Disinformation.” Policy Options, September 30, 2024. https://policyoptions.irpp.org/magazines/september-2024/defending-canada-the-battle-against-ai-driven-disinformation/.
Shan, Shawn, Wenxin Ding, Josephine Passananti, Stanley Wu, Haitao Zheng, and Ben Y. Zhao. “Nightshade: Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models.” arXiv, April 29, 2024. https://doi.org/10.48550/arXiv.2310.13828.
Shrivastava, Anika, Renu Rameshan, and Samar Agnihotri. “Latent Space Characterization of Autoencoder Variants.” arXiv, January 16, 2025. https://doi.org/10.48550/arXiv.2412.04755.
Simon, Felix M. “Artificial Intelligence in the News: How AI Retools, Rationalizes, and Reshapes Journalism and the Public Arena.” Tow Center for Digital Journalism: Columbia University, 2024. https://ora.ox.ac.uk/objects/uuid:aeb25013-1d17-40b2-b471-5bdca309db87.
Simon, Felix M., Sacha Altay, and Hugo Mercier. “Misinformation Reloaded? Fears about the Impact of Generative AI on Misinformation Are Overblown.” *Harvard Kennedy School Misinformation Review*, October 18, 2023. https://doi.org/10.37016/mr-2020-127.
Simon, Felix M, and Chico Q Camargo. “Autopsy of a Metaphor: The Origins, Use and Blind Spots of the

'Infodemic.'" *New Media & Society* 25, no. 8 (August 1, 2023): 2219–40. https://doi.org/10.1177/14614448211031908.
Simondon, Gilbert. *On the Mode of Existence of Technical Objects*. Univocal Publishing, 2017.
Sitaraman, Ganesh. "The Collapse of Neoliberalism." *The New Republic*, December 23, 2019. https://newrepublic.com/article/155970/collapse-neoliberalism.
Smuha, Nathalie A. "Beyond the Individual: Governing AI's Societal Harm." *Internet Policy Review* 10, no. 3 (September 30, 2021). https://policyreview.info/articles/analysis/beyond-individual-governing-ais-societal-harm.
Soares, Nate, and Benja Fallenstein. "Aligning Superintelligence with Human Interests: A Technical Research Agenda." Machine Intelligence Research Institute, 2014.
Solaiman, Irene, Zeerak Talat, William Agnew, Lama Ahmad, Dylan Baker, Su Lin Blodgett, Hal Daumé III, et al. "Evaluating the Social Impact of Generative AI Systems in Systems and Society." arXiv, June 12, 2023. https://doi.org/10.48550/arXiv.2306.05949.
Solomun, Sonja. "Polarization as the Technological Goal – Not the Error." Centre for Media, Technology and Democracy. Accessed August 6, 2024. https://www.mediatechdemocracy.com/all-work/polarization-as-the-technological-goal-not-the-error.
Solon, Olivia. "Facebook's Fake News: Mark Zuckerberg Rejects 'crazy Idea' That It Swayed Voters." *The Guardian*, November 11, 2016, sec. Technology. https://www.theguardian.com/technology/2016/nov/10/facebook-fake-news-us-election-mark-zuckerberg-donald-trump.
Sontag, Susan. *On Photography*. New York: Anchor Books, 1977.
Spataro, Jared. "Introducing Microsoft 365 Copilot – Your Copilot for Work." The Official Microsoft Blog, March 16, 2023. https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/.
Special Competitive Studies Project. "Defense Department Adoption of Generative Artificial Intelligence." Accessed May 29, 2024. https://www.scsp.ai/reports/gen-ai/defense/.
Srnicek, Nick. *Platform Capitalism*. John Wiley & Sons, 2017.
Staal, Jonas. "Assemblism - Journal #80." e-flux. Accessed May 23, 2025. https://www.e-flux.com/journal/80/100465/assemblism/.
———. *Propaganda Art in the 21st Century*. Cambridge, Massachusetts: The MIT Press, 2019.
Staniland, Martin. *What Is Political Economy?: A Study of Social Theory and Underdevelopment*. New Haven: Yale University Press, 1985.
State Council of China, *Notice of the Development Plan for the New Generation of Artificial Intelligence*, July 2017, https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm
———. *Notice of the State Council on Issuing 'Made in China 2025'*, Central People's Government of the People's Republic of China, May 2015, https://www.gov.cn/zhengce/content/2015-05/19/content_9784.htm
State Internet Information Office, *Global AI Governance Initiative*, Office of the Central Cyberspace Affairs Commission, October 2023, https://www.cac.gov.cn/2023-10/18/c_1699291032884978.htm
———. *Internet Information Service Algorithm Recommendation Management Regulations*, December 2021, https://www.gov.cn/zhengce/zhengceku/2022-01/04/content_5666429.htm
———. *Interim Measures for Generative Artificial Intelligence Service Management (draft),* April 2023, https://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm
———. *Interim Measures for Generative Artificial Intelligence Service Management*, July 2023, https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm
———. *Internet Service Deep Synthesis Management Provisions*, November 2022, https://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm
———. *Provisions on the Governance of the Online Information Content Ecosystem*, December 2019, https://www.cac.gov.cn/2019-12/20/c_1578375159509309.htm
Stephenson, Neal. *The Diamond Age*. New York: Bantam Books, 1995.
Sterling, Peter. "Allostasis: A Model of Predictive Regulation." *Physiology & Behavior* 106, no. 1 (April 12, 2012): 5–15. https://doi.org/10.1016/j.physbeh.2011.06.004.
Stiegler, Bernard. *For a New Critique of Political Economy*. 1st ed. Polity, 2010.
Stroebel, Leslie, and Richard D. Zakia. *The Focal Encyclopedia of Photography*. Focal Press, 1993.
Stone, Mike, and Aishwarya Jain. "Defense Firm Anduril Partners with OpenAI to Use AI in National Security Missions." Reuters, December 4, 2024. https://www.reuters.com/technology/artificial-intelligence/defense-firm-anduril-partners-with-openai-use-ai-national-security-missions-2024-12-04/.

Sturken, Marita, Douglas Thomas, and Sandra Ball-Rokeach. *Technological Visions: The Hopes and Fears That Shape New Technologies*. Temple University Press, 2004.

Suchman, Lucy. "Epilogue" in *Human-Machine Reconfigurations: Plans and Situated Actions*, 2nd ed. (New York: Cambridge University Press, 2007).

———. "Six Unexamined Premises Regarding Artificial Intelligence and National Security." *AI Now Institute* (blog), March 31, 2021. https://ainowinstitute.org/publication/six-unexamined-premises-regarding-artificial-intelligence-and-national-security.

Sun, Chen, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. "Revisiting Unreasonable Effectiveness of Data in Deep Learning Era." arXiv, August 4, 2017. https://doi.org/10.48550/arXiv.1707.02968.

Susskind, Jamie. "What Discourse Regulation by Social Media Giants Means For Democratic Societies." *Literary Hub* (blog), July 5, 2022. https://lithub.com/what-discourse-regulation-by-social-media-giants-means-for-democratic-societies/.

Sylvia Wynter: On Being Human as Praxis. Duke University Press, 2015. https://doi.org/10.2307/j.ctv11cw0rj.

Taffel, Sy. "Data and Oil: Metaphor, Materiality and Metabolic Rifts." *New Media & Society* 25, no. 5 (May 1, 2023): 980–98. https://doi.org/10.1177/14614448211017887.

Tagg, John. *The Burden of Representation: Essays on Photographies and Histories*. U of Minnesota Press, 1993.

Terrorist Content Analytics Platform. "Terrorist Content Analytics Platform." Accessed January 16, 2025. https://www.terrorismanalytics.org/.

Thaler, Richard H., and Cass R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, 2008.

"The Censorship Industrial Complex | United States Senate Committee on the Judiciary." Accessed April 9, 2025. https://www.judiciary.senate.gov/committee-activity/hearings/the-censorship-industrial-complex.

The Clear and Present Danger of Disinformation. World Economic Forum Annual Meeting. World Economic Forum, 2023. https://www.weforum.org/events/world-economic-forum-annual-meeting-2023/sessions/the-clear-and-present-danger-of-disinformation/.

The Economist. "China Is Writing the World's Technology Rules." October 10, 2024. https://www.economist.com/business/2024/10/10/china-is-writing-the-worlds-technology-rules.

"The Global Risks Report 2024." World Economic Forum, January 10, 2024. https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf.

The Globe and Mail. "Globe Editorial: Technology and the Law: Catching up to the Existential Threat of Deepfakes." June 27, 2024. https://www.theglobeandmail.com/opinion/editorials/article-technology-and-the-law-catching-up-to-the-existential-threat-of/.

The Government of Japan. "The Hiroshima AI Process: Leading the Global Challenge to Shape Inclusive Governance for Generative AI," February 9, 2024. https://www.japan.go.jp/kizuna/2024/02/hiroshima_ai_process.html.

The Markup. "About Us." Accessed March 14, 2025. https://themarkup.org/about.

Thompson, Adrienne. "The Existential Threat of AI-Enhanced Disinformation Operations." *Center for Security and Emerging Technology* (blog), July 11, 2022. https://cset.georgetown.edu/article/the-existential-threat-of-ai-enhanced-disinformation-operations/.

Traylor, Jake. "No Quick Fix: How OpenAI's DALL·E 2 Illustrated the Challenges of Bias in AI." NBC News, July 27, 2022. https://www.nbcnews.com/tech/tech-news/no-quick-fix-openais-dalle-2-illustrated-challenges-bias-ai-rcna39918.

"Trends in Internet Use and Attitudes: Findings from a Survey of Canadian Internet Users." The Strategic Counsel, March 2024. https://www.cira.ca/uploads/2024/07/2024-Internet-Trends-Factbook-FINAL-Report.pdf.

Trish, Barbara A. "4 Ways AI Can Be Used and Abused in the 2024 Election, from Deepfakes to Foreign Interference." The Conversation, October 16, 2024. http://theconversation.com/4-ways-ai-can-be-used-and-abused-in-the-2024-election-from-deepfakes-to-foreign-interference-239878.

TruePic. "C2PA Digital Content Provenance Virtual Event," Youtube, 2023, https://www.youtube.com/watch?v=vrrA_UJTZHM.

Tsui, Chin-Kuei. "The Securitization of Disinformation: Taiwan's Fake News Phenomenon and Anti-Disinformation Initiatives." *Tamkang Journal of International Affairs* 24, no. 2 (October 2020): 1–53.

Tubaro, Paola, Antonio A Casilli, and Marion Coville. "The Trainer, the Verifier, the Imitator: Three Ways in Which Human Platform Workers Support Artificial Intelligence." *Big Data & Society* 7, no. 1 (January 1, 2020): 2053951720919776. https://doi.org/10.1177/2053951720919776.

Tucker, Joshua A., Andrew Guess, Pablo Barbera, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. "Social Media, Political Polarization, and Political Disinformation: A Review

of the Scientific Literature." SSRN Scholarly Paper. Rochester, NY, March 19, 2018. https://doi.org/10.2139/ssrn.3144139.
Turing, Alan Mathison. *The Essential Turing*. Oxford University Press, 2004.
Tuters, Marc. "Fake News." *Krisis: Journal for Contemporary Philosophy* 38, no. 2 (2018): 59–61.
United Nations Development Programme. "A 'Super Year' for Elections." Accessed November 26, 2024. https://www.undp.org/super-year-elections.
United States Army Public Affairs. "Army Launches Detachment 201: Executive Innovation Corps to Drive Tech Transformation," June 13, 2025. https://www.army.mil/article/286317/army_launches_detachment_201_executive_innovation_corps_to_drive_tech_transformation.
United States Department of Commerce. "Implementation of Additional Export Controls: Certain Advanced Computing Items; Supercomputer and Semiconductor End Use; Updates and Corrections," Federal Registry 0694AI94, October 25, 2023. https://www.federalregister.gov/d/2023-23055
———. "Statement from U.S. Secretary of Commerce Howard Lutnick on Transforming the U.S. AI Safety Institute into the Pro-Innovation, Pro-Science U.S. Center for AI Standards and Innovation," June 3, 2025. https://www.commerce.gov/news/press-releases/2025/06/statement-us-secretary-commerce-howard-lutnick-transforming-us-ai.
United States Department of Defense. "Contracts for June 16, 2025," June 16, 2025. https://www.defense.gov/News/Contracts/Contract/Article/4218062//.
United States Deptartment of the Army. *The U.S. Army/Marine Corps Counterinsurgency Field Manual*. Chicago: The University of Chicago Press, 2007.
"U.S. Artificial Intelligence Safety Institute." National Institute of Standards and Technology, February 8, 2024. https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute.
Vaccari, Cristian, and Andrew Chadwick. "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News." *Social Media + Society* 6, no. 1 (January 1, 2020): 2056305120903408. https://doi.org/10.1177/2056305120903408.
Varoquaux, Gaël, Alexandra Sasha Luccioni, and Meredith Whittaker. "Hype, Sustainability, and the Price of the Bigger-Is-Better Paradigm in AI." arXiv, March 1, 2025. https://doi.org/10.48550/arXiv.2409.14160.
Venkatachary, Srinivasan. "A New Way to Search with Generative AI: An Overview of SGE." Alphabet, January 2024. https://static.googleusercontent.com/media/www.google.com/en//search/howsearchworks/google-about-SGE.pdf.
Virilio, Paul. *The Vision Machine*. Perspectives. Indiana University Press, 1994.
Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The Spread of True and False News Online." *Science* 359, no. 6380 (March 9, 2018): 1146–51. https://doi.org/10.1126/science.aap9559.
Vykopal, Ivan, Matúš Pikuliak, Ivan Srba, Robert Moro, Dominik Macko, and Maria Bielikova. "Disinformation Capabilities of Large Language Models." In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 14830–47, 2024. https://doi.org/10.18653/v1/2024.acl-long.793.
Walden, Scott. "Objectivity in Photography." *The British Journal of Aesthetics* 45, no. 3 (July 1, 2005): 258–72. https://doi.org/10.1093/aesthj/ayi036.
Wagner, Kurt. "Meta Opens Its AI Models to US Defense Agencies and Contractors." *Bloomberg.Com*, November 4, 2024. https://www.bloomberg.com/news/articles/2024-11-04/meta-opens-llama-ai-models-to-us-defense-agencies-contractors.
Wang, Pei. "On Defining Artificial Intelligence." *Journal of Artificial General Intelligence* 10, no. 2 (January 1, 2019): 1–37. https://doi.org/10.2478/jagi-2019-0002.
Wark, McKenzie. *Capital Is Dead: Is This Something Worse?* Verso Books, 2019.
Watkins, Ronald. "USSOCOM to Use AI to Detect Disinformation Threats on Social Media." The Defense Post, August 31, 2023. https://www.thedefensepost.com/2023/08/31/ussocom-ai-disinformation-social-media/.
Weatherbed, Jess. "Meta Fed Its AI on Almost Everything You've Posted Publicly since 2007." The Verge, September 12, 2024. https://www.theverge.com/2024/9/12/24242789/meta-training-ai-models-facebook-instagram-photo-post-data.
Weatherby, Leif. "ChatGPT Is an Ideology Machine." Jacobin, April 17, 2023. https://jacobin.com/2023/04/chatgpt-ai-language-models-ideology-media-production.
Weber, Valentin. "Data-Centric Authoritarianism: How China's Development of Frontier Technologies Could Globalize Repression." Washington, D.C: National Endowment for Democracy, February 11, 2025. https://www.ned.org/data-centric-authoritarianism-how-chinas-development-of-frontier-technologies-

could-globalize-repression-2/.
Wei, Jason, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, et al. "Emergent Abilities of Large Language Models." arXiv, October 26, 2022. https://doi.org/10.48550/arXiv.2206.07682.
Wei, Lu. "National Discourse Power and Information Security Against the Background of Economic Globalization." Translated by Rogier Creemers. DigiChina - Stanford Cyber Policy Center, December 21, 2014. https://digichina.stanford.edu/work/national-discourse-power-and-information-security-against-the-background-of-economic-globalization/.
Welbl, Johannes, Amelia Glaese, Jonathan Uesato, Sumanth Dathathri, John Mellor, Lisa Anne Hendricks, Kirsty Anderson, Pushmeet Kohli, Ben Coppin, and Po-Sen Huang. "Challenges in Detoxifying Language Models." In *Findings of the Association for Computational Linguistics: EMNLP 2021*, edited by Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, 2447–69. Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021. https://doi.org/10.18653/v1/2021.findings-emnlp.210.
West, Sarah Myers. "Data Capitalism: Redefining the Logics of Surveillance and Privacy." *Business & Society* 58, no. 1 (January 1, 2019): 20–41. https://doi.org/10.1177/0007650317718185.
Whittaker, Meredith. "The Steep Cost of Capture." SSRN Scholarly Paper. Rochester, NY, 2021. https://papers.ssrn.com/abstract=4135581.
Williams, Alex. "Control Societies and Platform Logic. ('Postscript on the Societies of Control' by Gilles Deleuze)." *New Formations*, no. 84–85 (2015). https://doi.org/10.398/NEWF:84/85.10.2015.
Williams, Dan. "There Is No 'Censorship Industrial Complex.'" *Conspicuous Cognition* (blog), July 28, 2024. https://www.conspicuouscognition.com/p/there-is-no-censorship-industrial.
Williams, Kristian, William Munger, and Lara Messersmith. *Life during Wartime: Resisting Counterinsurgency*. Oakland, CA, USA: AK Press, 2013.
Wilson, Japhy, and E. Swyngedouw. *The Post-Political and Its Discontents: Spaces of Depoliticisation, Spectres of Radical Politics*. Edinburgh: University Press, 2014.
Winner, Langdon. "Do Artifacts Have Politics?" *Daedalus* 109, no. 1 (1980): 121–36.
Wolfe, Frank. "San Francisco-Based Firm Announces DoD SBIR Contracts for Deep Fake Detection." Defense Daily, November 28, 2022. https://www.defensedaily.com/san-francisco-based-firm-announces-dod-sbir-contracts-for-deep-fake-detection/air-force/.
Wolin, Sheldon S. *Democracy Incorporated: Managed Democracy and the Specter of Inverted Totalitarianism - New Edition*. Princeton University Press, 2017.
X Safety. "Freedom of Speech, Not Reach: An Update on Our Enforcement Philosophy." Blog, April 17, 2023. https://blog.x.com/en_us/topics/product/2023/freedom-of-speech-not-reach-an-update-on-our-enforcement-philosophy.
Ye, Hongbin, Tong Liu, Aijia Zhang, Wei Hua, and Weiqiang Jia. "Cognitive Mirage: A Review of Hallucinations in Large Language Models." arXiv, September 13, 2023. https://doi.org/10.48550/arXiv.2309.06794.
Yeung, Karen. "'Hypernudge': Big Data as a Mode of Regulation by Design." *Information, Communication & Society* 20, no. 1 (January 2, 2017): 118–36. https://doi.org/10.1080/1369118X.2016.1186713.
Yu, Zhiyuan, Xiaogeng Liu, Shunning Liang, Zach Cameron, Chaowei Xiao, and Ning Zhang. "Don't Listen To Me: Understanding and Exploring Jailbreak Prompts of Large Language Models." arXiv, March 25, 2024. https://doi.org/10.48550/arXiv.2403.17336.
Zerback, Thomas, Florian Töpfl, and Maria Knöpfle. "The Disconcerting Potential of Online Disinformation: Persuasive Effects of Astroturfing Comments and Three Strategies for Inoculation against Them." *New Media & Society* 23, no. 5 (May 1, 2021): 1080–98. https://doi.org/10.1177/1461444820908530.
Zhang, Xia-heng. "Political and Social Dynamics, Risks, and Prevention of ChatGPT," *Journal of Shenzhen University (Humanities & Social Sciences),* 40, no. 3 (2023).
Zirpoli, Christopher. "Generative Artificial Intelligence and Copyright Law." *Copyright, Fair Use, Scholarly Communication, Etc.*, February 24, 2023. https://digitalcommons.unl.edu/scholcom/243.
Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, 2019.