Predicting Business Angel Early-Stage Decision Making Using AI

Yan Katcharovski

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

GRADUATE PROGRAM IN MECHANICAL ENGINEERING
YORK UNIVERSITY
TORONTO, ONTARIO

January 2026

Abstract

External funding is crucial for early-stage ventures, yet business angel decision-making remains subjective and resource-intensive. The Critical Factor Assessment (CFA), a validated eight-factor venture evaluation framework, has demonstrated superior predictive accuracy over investors' own decisions. However, full evaluation requires multiple trained evaluators and several days per assessment, limiting adoption at scale. This study investigates whether AI can overcome these constraints. Multiple Large Language Models (LLMs) were prompted to assign CFA scores to 600 transcribed Shark Tank pitches with known deal outcomes. Machine learning classification models trained on the LLM-generated CFA scores achieved 85.0% accuracy in predicting deal/no-deal outcomes. The top-performing model (GPT-4.1-mini) exhibited very strong correlation with trained human evaluators (Spearman's $\rho = 0.909$, $p < .001$), substantially exceeding mean human–human agreement ($\rho = 0.465$). The integration of AI-based feature extraction with a validated decision-making framework yielded a scalable, reliable approach to early-stage venture evaluation.

Author's Declaration

I hereby declare that this thesis was produced through my sole efforts and supervised by Dr. Andrew Maxwell during my master's research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

1.1 Background and Context

Startup formation rates have hit historical records during the last two decades, continuing to establish itself as a fundamental driver of economic development (Kane, 2010). The US alone has seen a 50% increase since 2019 levels, with the United States Department of Treasury reporting an average of 430,000 new businesses registered per month in 2024 (Nostrand, 2025). Entrepreneurial activities drive economic growth through innovation, making way for technological development, employment, and wealth (Kritikos, 2015). Among the various important aspects of entrepreneurship and venture creation, the vital struggle for early-stage venture funding stands as a central topic of interest and concern. Early-stage ventures lack the characteristics of established businesses, such as track record and consumer validation, thereby presenting investors with investment opportunities that are inherently uncertain, all while necessitating product development costs prior to stable revenue generation (Rasmussen & Sørheim, 2012). The crucial funding period, known as the "valley of death", is particularly intensified for technology-based startups as they must spend significant funds on innovative product research and development and market entry prior to becoming profitable (Harris, 2013). Business Angels (BAs) serve as the essential early-stage investors who bridge the gap between market validation and institutional venture capital in this funding ecosystem. Venture capitalists operate through pooled funds with fiduciary duties toward limited partners, whereas BAs use their personal wealth to invest directly in startups through individual investment choices, based on their assessment of opportunity risk and potential return (Freear et al., 1994). As a result, BAs gain flexibility to take higher risks and deliver more than financial support through mentorship, advisory services, and valuable industry connections. Furthermore, BA-backed businesses achieve better growth performance while facing lower death rates and securing institutional funding more frequently than ventures supported only by bootstrapping (no funding from external sources) or informal funding (Freear et al., 1994; Macht & Robinson, 2009). The

2023 U.S. angel investment market delivered $18.6 billion to startups through 54,735 deals, acting as the primary funding mechanism for seed-stage companies (Sohl, 2023). The BA investment process, while crucial for the early-stage startup ecosystem, is riddled with complexity and challenges, thereby reducing both its operational performance and effectiveness. The combination of investment decision subjectivity through human judgment flaws and cognitive biases generates systematic inefficiencies that negatively affect both investors and entrepreneurs (Baker & Nofsinger, 2002).

Research demonstrates that BA investment decisions remain highly subjective and inconsistent. Individual investors often evaluate the same venture pitch differently, even though possessing similar backgrounds and experience (Armandi, 2015; Madaan & Singh, 2019; Murnieks et al., 2011). The inconsistent assessments resulting from human judgment errors and mental shortcuts are further exacerbated by the rapidly changing and uncertain market dynamics, especially in the age of artificial intelligence. The thorough evaluation process requires substantial resources which makes it challenging for both detailed due diligence and quality assessment accessibility. Even in professional structured evaluation settings, experienced professionals must spend considerable time using traditional evaluation methods which restricts the number of ventures that can be filtered and assessed, further making capital more expensive for entrepreneurs (Fisher & Neubert, 2022; Klonowski, 2018). The lack of resources becomes further problematic as the number of startups actively seeking angel funding is growing. Investors and entrepreneurs operating in different ecosystems lack standardized evaluation frameworks, preventing entrepreneurs from receiving consistent feedback that would help understand investor decision-making and develop their ventures effectively (Olaore & Adegbesan, 2014; Steier & Greenwood, 1999). The result: The funding process becomes less efficient while the potential for education and improvement from investor feedback decreases due to information asymmetry. Existing decision support tools have proven their effectiveness, but their practical adoption remains restricted due to high implementation expenses and demanding training for implementation and scale (Maxwell et al., 2011). The absence of systematic evaluation tools further reinforces biases that negatively affect

specific underrepresented entrepreneurial groups and business types, thus restricting diversity and inclusiveness within innovation ecosystems (Fischer et al., 2023).

In efforts to rectify some of the challenges, decision support systems have been developed by academics and entrepreneurial ecosystem stakeholders with the purpose of improving the quality, efficiency, and consistency of early-stage investment evaluation.

While the validated critical factor assessment (CFA) offers a reliable evaluation framework that can be used to provide specific and relevant guidance to the entrepreneur, its use in practice has been hampered by the time-consuming and resource-intensive characteristics of its deployment (Maxwell et al., 2011).

This study builds on previous research on the CFA framework, widely used and validated in entrepreneur development and investment assessments, to develop an AI operationalization model. We addressed the challenges of widespread deployment and scalability of the original tool, given the previously identified implementation constraints of cost, training, and potential biases (Maxwell et al., 2011). The results of this research reinforce and replicate the original deployment of the CFA tool using the same eight factors as in Andrew Maxwell's original research: Features & Benefits, Readiness, Market Size, Barriers to Entry, Supply Chain, Entrepreneurial Experience, Financial Expectations, and Adoption (Maxwell et al., 2011). This study showcases how well-designed and trained AI tools can provide valuable insights for entrepreneurs and those working with them to identify specific venture challenges that must be addressed to increase the likelihood of receiving funding.

This study was facilitated by our access to the original data and assessment techniques used in prior work (Maxwell et al., 2011), knowledge of the CFA process, advancements in large language models (LLMs), and supervised machine learning algorithms. This allowed us to automate the evaluation of startup pitches, synthesize factors that impact BA decisions, and provide diagnostic feedback to the entrepreneur. The dataset included transcripts from 600 pitches presented on the U.S. television show Shark Tank, providing a diverse sample of real-world interactions between investors and

entrepreneurs, resembling those in conventional funding settings. The validity of this dataset to replicate real decision-making was discussed in Maxwell's research that studied full, behind-the-scenes pitches from the Canadian television show Dragon's Den (briefly addressed in this paper) (Maxwell et al., 2011).

1.2 Research Objectives and Questions

The research aims to create and assess an AI-enhanced version of the Critical Factor Assessment (CFA) framework which preserves its traditional human-administered accuracy while addressing barriers to adoption. The main research inquiry investigates whether Large Language Models can duplicate trained human evaluation precision when using the validated Critical Factor Assessment for early-stage venture pitch assessment.

Research Questions:

1. When evaluating startup pitches, would a CFA-augmented AI model produce consistent and accurate CFA evaluations across each of the eight critical factors when compared to the scoring of trained human evaluators?
2. When forecasting an investment decision of a group of BAs, will a CFA-augmented AI model accurately predict the outcome?
3. Would a CFA-prompted LLM produce better evaluations and investment predictions than an unprompted "stock" LLM (i.e., out-of-the-box GPT-4)?
4. Would a CFA-prompted LLM would produce better and more relevant feedback and advice to the entrepreneur than an unprompted "stock" LLM?

Research Objectives:

1. A validated AI system would be developed to produce CFA evaluations from unstructured venture pitch transcripts.

2. Investigate how closely AI-generated CFA evaluations match human expert predictions by using established correlation and classification metrics.
3. A comparison would be made between framework-guided and generic AI approaches for venture evaluation.
4. To identify the most predictive individual CFA factors and their relative importance in AI-based assessment.
5. To study and analyze the real-world applications and implications of using AI-enhanced CFA for all parties in the entrepreneurial ecosystem.

1.3 Significance of the Study

This research contributed insights into the areas of entrepreneurship theory, investment practice, and the application of artificial intelligence to investment decisions. In this study, we developed a theoretical understanding of how artificial intelligence can improve or augment human decision-making processes in difficult and uncertain situations that require analysis of unstructured data. This research showed that AI can effectively leverage established decision-making frameworks for unstructured business data, expanding the body of knowledge that examines the conditions for successful human-AI teamwork and technology-driven enhancement of human expertise. Furthermore, this study introduced a new approach to previous research that shows how AI can support proven human decision-making systems, rather than focusing on replacing decision-making with algorithms. Finally, entrepreneurs gain access to a practical way to receive structured feedback on the development of their business. Through an automated CFA assessment, entrepreneurs can receive a standardized assessment of their venture viability, allowing them to focus their attention on the "correct" improvements and improve their preparedness for investor interactions and market viability. The democratization of validated assessments creates significant value for entrepreneurs who have limited access to professional advisors or experienced angel investors due to their geography, networks, funding, or other reasons. Additionally, AI-augmented CFA offers

investors an affordable, low-risk way to screen early-stage ventures according to standardized criteria that investors already use to evaluate ventures. The tool can be deployed as a decision support tool using established evaluation dimensions to highlight the key strengths and weaknesses of a venture investment opportunity, allowing investors to allocate their time efficiently (e.g. to perform detailed due diligence instead). This study demonstrated that an academic assessment framework can be an operational tool to support the development of an entrepreneurial ecosystem. AI-augmented CFA systems can be implemented in entrepreneurship education programs, accelerator curricula, and startup support services to deliver standardized assessment capabilities that require expert-level knowledge. This study also contributed to explainable AI research by showing how established decision-making frameworks can improve the interpretability of AI evaluation results. The CFA-augmented approach links evaluation outcomes directly to assessment criteria, providing the clarity of applicability needed for transparency requirements in high-stakes AI applications.

1.4 Scope and Delimitations

The research investigated early-stage venture evaluation through the eight-factor Critical Factor Assessment framework created by Andrew Maxwell (Maxwell et al., 2011). The study used only English-language Shark Tank pitches of ventures seeking funding.

Scope Inclusions:

- LLM-based factor (feature) assignment using the validated CFA framework across all eight assessment dimensions
- Analysis of LLM-generated evaluations through established correlation and predictive accuracy metrics compared to trained human assessments
- Investigation of assessment factor importance for AI systems and their relative predictive strength
- Evaluation of framework-based LLM assessment approach performance against unprompted, out-of-the-box LLMs

- Evaluation of practical applications for different entrepreneurial-ecosystem stakeholders

Delimitations:

- The research did not investigate how investments perform after funding decisions or how ventures performed in the long term beyond their initial funding stage.
- The research investigated the CFA framework exclusively without evaluating alternative validated assessment methods.
- The research utilized Large Language Models offered by OpenAI and did not investigate alternative LLM providers (e.g. Google, Anthropic, Meta).
- The research focused on ventures from the United States, which may reduce its general applicability to entrepreneurial environments outside this context.

1.5 Key Definitions

- **Business Angels (BAs):** Private investors who invest their personal funds into start-ups at their early stages, often providing entrepreneurial expertise as well as market and financial assistance.
- **Early-stage ventures:** Nascent companies that have no established history and uncertain market prospects that often need significant funding before they can reach profitability.
- **The Valley of Death:** The funding gap which exists between high R&D costs and sufficient revenue generation.
- **Critical Factor Assessment (CFA):** A validated eight-factor framework for evaluating early-stage venture viability, originally developed by the Canadian Innovation Centre and refined by Andrew Maxwell to provide systematic assessment criteria for investment decision-making.
- **The eight CFA factors:** Features & Benefits (product/service value proposition and competitive advantage); Readiness (market entry preparedness and

operational capabilities); Market Size (addressable market potential and growth prospects); Barriers to Entry (competitive protection and entry obstacles); Supply Chain (operational partnerships and distribution capabilities); Entrepreneurial Experience (founding team expertise and track record); Financial Expectations (realistic projections and capital requirements); Adoption (market acceptance potential and customer validation evidence) (Maxwell et al., 2011).

- **Non-compensatory decision-making:** The elimination-by-aspects methodology of decision-making eliminates opportunities when any critical factor performs poorly even when other factors show strength.
- **Artificial Intelligence (AI):** Technology that is capable of replicating human intelligence, thinking, problem solving, learning, and decision making.
- **Large Language Model (LLM):** Advanced artificial intelligence systems trained on vast text that can understand, process, and generate human-like responses.
- **Prompt engineering:** The systematic design and optimization of input instructions for LLM models to guide their information processing toward specific analytical frameworks, evaluation criteria, and output formats to achieve consistent and accurate results.
- **Feature extraction:** The process of identifying and converting relevant characteristics from unstructured data sources (such as text) into structured, quantifiable variables (such as CFA grades, later codified) suitable for machine learning analysis and model training.
- **Machine-learning classification model:** Algorithms that learn patterns from training data to categorize new inputs into predefined classes or outcomes, such as predicting whether an investment opportunity will result in a deal or no-deal decision.
- **Predictive accuracy:** The proportion of correct predictions made by a model when tested on new, unseen data, representing the model's ability to generalize beyond its training dataset.
- **Model metrics:** Performance evaluation measures including F1 score (harmonic mean of precision and recall), recall (proportion of actual positives correctly

identified), precision (proportion of predicted positives that are correct), specificity (proportion of actual negatives correctly identified), and confusion matrix (tabular display of prediction accuracy across all categories).

- **False positive / false negative errors:** Classification mistakes where false positives incorrectly predict success (i.e., a deal outcome) when failure occurs (Type I error) and false negatives incorrectly predict failure when success occurs (Type II error), each carrying different costs in investment contexts.
- **Cross-validation:** A statistical technique for assessing model performance by repeatedly dividing data into training and testing subsets, ensuring reliable performance estimates while maximizing data use and preventing overfitting to specific data patterns.
- **Holdout test set:** The holdout test set consists of 10-20% of the dataset which stays untouched during model training and hyperparameter tuning to serve as an evaluation metric.
- **Inter-rater reliability:** The extent of agreement between multiple evaluators who assess identical subjects using the same criteria.
- **Ensemble methods:** Machine learning approaches that merge predictions from multiple single models to generate better results than any model operating alone.
- **Hyperparameter tuning:** The optimization of model configuration parameters to achieve maximum validation data performance by adjusting learning rates, regularization parameters, and algorithm-specific options.
- **Algorithmic bias:** The output of AI systems exhibits systematic errors and unfair discrimination because of training data bias along with model design flaws and implementation choices.

1.6 Thesis Structure

This thesis is organized into eight chapters: literature review, methodology, results, and discussion.

Chapter 1 provides an introduction for the background, motivation, approach and structure of the research.

Chapter 2 provides a comprehensive literature review covering business angel decision-making, artificial intelligence applications in finance, and the theoretical foundations of decision support systems.

Chapter 3 establishes the theoretical framework that guides the research approach and hypothesis development.

Chapter 4 details the research methodology, including data collection procedures, AI model development, and evaluation metrics.

Chapter 5 presents the empirical results, including performance comparisons and factor analysis.

Chapter 6 provides extensive discussion of findings, implications, and connections to existing literature.

Chapter 7 addresses research limitations and outlines future research opportunities

Chapter 8 concludes with a synthesis of contributions and final recommendations for researchers, practitioners, and policy makers.

Supporting materials, including detailed sample pitches, outputs, prompts, methodologies and analyses, are provided in the appendices.

Chapter 2: Literature Review

2.1 Business Angel Investment Decision-Making

2.1.1 Behavioral Decision Theory and Decision Biases

The foundation for analyzing investment decisions originates from behavioral decision theory, developed in the intersection of psychology and economics research. The field of behavioral decision theory, popularized by Herbert Simon, Daniel Kahneman, and Amos Tversky, proved that pure economic rationality does not exist; the research found various judgment and choice patterns which differ from optimal decision methods (Kahneman, 2003). The study of investment decision-making is largely informed and influenced by behavioral decision theory, specifically as it pertains to how investors process information and form judgments about uncertain outcomes (especially while operating with limited information and time constraints) (Kahneman, 2003). The dual-process theory developed by Kahneman demonstrated that human thinking operates through two separate systems: System 1 represents fast, automatic, and intuitive thinking. System 2 represents slow, deliberate, and analytical thinking (Kahneman, 2011). Early-stage venture assessment typically takes place in time sensitive and information lacking environments, thereby increasing the probability of System 1 processing. This results in decisions that depend on heuristics and biases instead of analytical methods.

The availability heuristic described by Tversky and Kahneman directly applies to the evaluation of ventures (Tversky & Kahneman, 1973). Investors tend to evaluate new opportunities through the lens of easily remembered past examples of successful or failed ventures, resulting in systematic errors when assessing risks and opportunities. Investors are also prone to the representativeness heuristic, tending to focus on superficial similarities between new ventures and successful familiar companies while disregarding fundamental differences in market conditions, execution capabilities, and competitive dynamics (Tversky & Kahneman, 1974).

Furthermore, entrepreneurs typically present their most optimistic projections first, thereby creating anchoring effects that distort subsequent judgments (Wiltbank et al., 2009). Angels display confirmation bias when they search for evidence which supports their first impressions while ignoring contradictory information. During due diligence, the process of systematic information collection becomes essential for accurate assessment, thus making this bias particularly problematic (Cheng, 2018).

2.1.2 Information Processing Theory in High-Uncertainty Contexts

The cognitive psychological origins of Information Processing Theory serve as an additional theoretical base to study venture evaluation challenges. According to this theory, human beings demonstrate restricted abilities in processing intricate ambiguous data, especially when experiencing time constraints and cognitive strain (resembling typical investment decision scenarios) (Deci & Ryan, 1985; Payne & Bettman, 2008; Stiensmeier-Pelster & Schürmann, 1993).

During venture evaluation processes, decision-makers need to combine financial projections with market assessments, team evaluations, technology assessments, and competitive analyses while dealing with high uncertainty regarding future results (Maxwell et al., 2011). According to the theory, human decision-makers employ a decision-making strategy called satisficing, by choosing the first acceptable option that passes a decision threshold (rather than optimizing and chunking information into manageable units, paired with sequential processing by addressing decision dimensions one at a time, according to classical information processing theory) (Caplin et al., 2011; Simon, 1956). These adaptive strategies result in suboptimal venture evaluation results. The practice of satisficing causes investors to accept the first available investment opportunity without performing full comparisons between alternatives. Investors who employ chunking tend to concentrate on familiar elements of ventures but disregard important yet unfamiliar factors. The method of sequential processing creates inconsistent evaluation criteria and weighting patterns when applied to different investment possibilities (Caplin et al., 2011). Modern information processing theory

integrates dual-process models to demonstrate that complex environment decisions need both pattern recognition capabilities and systematic analytical evaluation (Kahneman, 2011).

2.1.3 Decision Quality and Effectiveness Frameworks

To assess the effectiveness of decision support systems, it is necessary to establish specific methods for measuring decision quality. Traditional assessment of decision quality has focused on outcome measurements by evaluating actual results against predicted outcomes alongside optimal benchmarks. Outcome-based measurement methods create challenges for venture investment assessments because venture success outcomes span extended timeframes and are affected by a variety of internal and external factors beyond initial venture quality.

The study of the factors contributing to decision making rather than outcome results in process-based evaluation methods have been proposed (Arvai & Froschauer, 2010). The decision-making process may be better informed when it accounts and evaluates information systematically, using agreed-upon criteria that remain constant throughout the process (and subsequent evaluation processes of the same kind) (Hammond et al., 1998). According to Hammond, decision quality includes both process elements and outcome elements within a comprehensive framework:

- Value clarity means that all team members have a precise understanding of their objectives alongside the trade-offs involved
- Comprehensive consideration of all available options, including weighing relevant and reliable data
- Sound reasoning which excludes cognitive biases
- Commitment to execute its decisions through a unified framework.

The implementation and consideration of such characteristics of better decision quality within an evaluation framework would yield improved decision-making outcomes and reduced ambiguity (Hammond et al., 1998).

2.1.4 Evolution of Formal Business Angel Research

During the early years of study, Freear, Sohl, and Wetzel (Freear et al., 2002) conducted foundational investigations into angel demographics, investment value, and fundamental selection factors. Early research on business angels identified this investor group as distinct from both venture capitalists and informal backers (including friends and family members) (Prowse, 1998).

Researchers began analyzing decision-making processes rather than solely focusing on decision outcomes in business angel research. Harrison et al. (Harrison et al., 2015) established important process research through surveys studying how business angels assess investment opportunities. Research on self-reported investor decision-making data faced problems, with subsequent studies revealing that respondents tended to remember and justify their answers inaccurately (Harrison et al., 2015).

Wiltbank et al. (Wiltbank et al., 2009) employed conjoint analysis to determine which evaluation factors angels value most. The modern business angel research uses behavioral economics together with cognitive psychology to study the mental processes which guide investment choices.

Research conducted by Maxwell et al. (Maxwell et al., 2011), Mitteness et al. (Mitteness et al., 2012), and Brooks et al. (Brooks et al., 2014) used complex research methods to study how cognitive biases, heuristic decision-making, and emotional responses affect investment evaluation.

2.1.5 Decision-Making Processes and Criteria

It has been proposed that BA investment decisions proceed through a standard multiple-stage sequence, despite that fact that each investor draws from a unique and differing history and context. BAs perform rapid assessments during their initial screening phase to eliminate opportunities that do not satisfy their basic investment criteria or personal interests (Maxwell et al., 2011). According to Maxwell et al.

(Maxwell et al., 2011) BAs use a systematic elimination-by-aspects method to first identify fatal flaws that eliminate investment opportunities regardless of other positive aspects. After the initial screening process, angels proceed to successive evaluation stages starting with market analysis followed by financial due diligence and team evaluation, finishing with negotiation and structuring (Maxwell et al., 2011).

Research shows that angels evaluate their investment opportunities based on defined fundamental criteria during the entire decision process, with the entrepreneurial team standing as the leading evaluation factor for angels who prioritize relevant industry experience, track record of success and leadership capabilities and coachability (Mason & Stark, 2004; Sudek, 2006; Wiltbank et al., 2009). Angels prefer backing teams instead of opportunities because they believe talented entrepreneurs can turn average business ideas into successful ventures.

Angels assess market size alongside growth potential for investments while paying close attention to ventures which can reach significant scale within feasible time periods. The research shows that business angels struggle with market analysis because they heavily depend on information provided by entrepreneurs, especially in niche, new, and unfamiliar business categories (Mitteness et al., 2012).

Angels evaluate the technical feasibility and market appeal of venture core offerings by examining competitive advantages together with intellectual property protection measures. Business angels face challenges when evaluating advanced technical products that operate outside their areas of expertise (Macht, 2011). BAs examine financial projections, with the focus on evaluating the reasonableness of the basic assumptions rather than exact numerical accuracy (Mason & Stark, 2004).

2.1.6 Due Diligence Practices and Challenges

Institutional venture capitalists conduct extensive multiple-week evaluation processes with specialized professionals, yet angels perform fewer formal assessments that potentially miss crucial risks (Wong, 2002). Angel investors perform shorter

unsystematic reviews which primarily depend on personal instincts rather than those of venture capitalists. Rapid decision process and decreased transaction costs are notable benefits, though risk producing substandard risk evaluation and investment results (Paul et al., 2007).

BAs execute due diligence differently from one another according to research, which shows large discrepancies between optimal procedures and their real-world practices (Paul et al., 2007). Key challenges in angel due diligence include the lack of access to specialized analysts, technical experts, and industry specialists who can perform comprehensive assessments of complex technology or market dynamics (Gompers, 1995). Most BAs dedicate only part-time hours to investing as they maintain other professional obligations, thus restricting due diligence process by time limitations (Sørheim, 2003). Angels who focus on local investments can monitor their portfolio companies, but their geographical focus restricts their ability to seek expert evaluation for ventures operating in unfamiliar sectors or locations (Sørheim, 2003).

The assessment of business angel decision-making must consider their distinct methods compared to venture capital practices at both decision-making stages and results. The evaluation of fundamental criteria including team, market, product, and financial potential occurs at both angel and venture capital levels, though each investor type implements these criteria differently. Evaluation processes of venture capitalists include structured stages and formal documentation standards along with committee-based decision-making procedures (Gompers, 1995). The methodical evaluation process decreases personal cognitive errors yet produces herding behavior that drives multiple companies to follow the same investment targets while missing distinctive valuable opportunities (Scharfstein & Stein, 1990; Wong, 2002). Individual angel investors evaluate opportunities on their own while showing flexibility in their process, though assessment quality varies widely. According to Wong (Wong, 2002), BAs identify investments that institutional investors would not pursue due to size, stage, or strategic considerations. Angel investment and venture capital financing show growing

interconnectedness since many startups use angel capital to progress toward institutional financing. Angel investors must evaluate both present return opportunities and institutional funding potential during their decision-making process (Wong, 2002).

2.2 The Critical Factor Assessment

2.2.1 Historical Development and Theoretical Foundation

The Critical Factor Assessment (CFA) framework, developed in the 1980s by the Canadian Innovation Centre, is a structured method to validate early-stage venture commercialization viability (Astebro, 2004). Traditional financial analysis methods failed to meet the requirements for assessing ventures as they lack operating history, had uncertain and proven market potential, and evolving and untested business models (Astebro, 2004). The CFA framework uses theoretical foundations that combine strategic management with entrepreneurship theory and decision science principles.

The framework uses non-compensatory decision-making principles, rooted in Tversky's elimination-by-aspects decision theory. Tversky showed that a single critical weakness can eliminate investment possibilities regardless of other strengths (Tversky, 1972). The original, 37-factor CFA, developed through investor, entrepreneur, and expert industry consultation was further refined and validated in studies during the 1980s and 1990s (Astebro, 2004). The framework's development process incorporated both new entrepreneurship research theories and practical feedback obtained from real-world investment evaluation situations. The eight-factor CFA framework developed by Dr. Andrew Maxwell (which refined the 37 factors down to 8) achieved major theoretical and practical validation through both empirical and usability studies. The research findings by Maxwell confirmed that the decreased number of assessment factors preserved predictive performance while making evaluation process simpler less resource intensive (Maxwell et al., 2011).

2.2.2 The Eight-Factor Framework Structure

The eight-factor CFA framework provides a well-rounded method for early-stage venture assessment, including elements of market opportunity analysis together with execution capability evaluation. Research-based assessment criteria in the framework evaluate distinct venture viability factors which prove essential for early-stage success (Maxwell et al., 2011):

The Features & Benefits factor evaluates how well the venture's product or service delivers value to customers through performance improvements relative to competitors, cost-effectiveness, and customer requirement alignment. According to strategic management principles, sustainable competitive advantage requires distinctive customer value creation.

The Readiness factor evaluates how ready the venture is to enter the market through assessments of product development status, regulatory approvals, operational capabilities, and milestone achievements. The factor shows that excellent opportunities become ineffective when execution abilities are insufficient or market timing is incorrect.

The Market Size factor assesses the revenue potential through assessments of market addressability along with projected growth and penetration estimates. Investments need to access extensive markets to demonstrate growth potential for generating substantial value and financial return.

The Barriers to Entry factor examines how well the venture can protect its competitive advantages through the use of intellectual property protection, technical complexity, regulatory requirements, and other structural factors that create entry barriers.

The Supply Chain factor evaluates the venture's operational readiness, including supplier relationships, distribution capabilities, manufacturing capacity, and other execution-critical partnerships and capabilities.

The Entrepreneurial Experience factor evaluates the founding team's experience levels and their execution capabilities together with their knowledge of the industry and past achievements.

The Financial Expectations factor evaluates whether financial projections together with capital needs and return targets align with current market situations, feasibility, and venture attributes.

The Adoption factor evaluates market acceptance potential by examining customer validation, early adoption indicators, sales traction, and product-market fit evidence.

Between 1976 and 2004, the Canadian Innovation Assistance Program (IAP), in collaboration with the CIC, used the CFA to evaluate over 14,000 innovative venture applications (Maxwell, 2009). Empirical studies on Maxwell's CFA framework have shown its efficacy in structuring BA decision-making. Investor behavior analysis in quasi-experimental settings, exemplified by the Canadian reality television show Dragons' Den, further validated that BAs utilize elimination-by-aspects (EBA) heuristics, leading to the immediate rejection of opportunities with critical flaws (Maxwell et al., 2011. The CFA framework yielded a predictive accuracy of 87.5% when forecasting which ventures progress beyond the initial selection stage, as BAs consistently rejected pitches that score poorly on critical factors (Maxwell et al., 2011). The CFA was also found to be an accurate predictor of whether early-stage ventures would go on to succeed at go-to-market product launch. A study that followed 561 CFA-screened ventures five years post-assessment found that the critical factors predicted ventures that were still in business with an accuracy of 80.9% (Astebro, 2004).

2.3 Artificial Intelligence in Investment Analysis

2.3.1 Evolution from Rule-Based to Machine Learning to Large Language Models

Artificial intelligence applications in investment analysis have developed through three major technological phases. During the rule-based era spanning from the 1980s to the early 1990s, expert knowledge was translated and codified into decision trees and

logical rules for standardized investment opportunity assessment. Rule-based investment systems developed by financial institutions during the early days showed potential for automated decision support, yet faced challenges related to system maintenance and brittleness (in its ability to generalize to new data and adapt to a changing market) (Trippi & Lee, 1995).

The machine learning era emerged in the 1990s until the 2010s with focus on data analysis for pattern recognition and predictive modeling. Machine learning techniques such as neural networks, support vector machines, and random forests allowed the discovery of patterns from historical data without requiring predefined rules (Dietterich, 1996; Jordan & Mitchell, 2015). The quantitative investment methods proved especially effective when dealing with structured data sources such as market data patterns and financial metrics (Lai & Xing, 2008). The processing capabilities of traditional machine learning systems did not perform as well in the analysis of unstructured data, typically found in analyst text-based reports, news articles, and qualitative information needed for thorough investment analysis. The processing of textual data required advanced feature engineering along with extensive preprocessing to obtain meaningful insights (Crowston et al., 2012; Schumaker & Chen, 2006).

The development of Large Language Models (LLMs) during the late 2010s introduced transformative abilities for handling and analyzing unstructured data. GPT-3 and GPT-4 alongside their derivatives demonstrate abilities to interpret intricate business ideas, interpret narrative business plans, and qualitative elements that previously relied on human assessment (Chiang & Lee, 2023; Liu et al., 2023).

2.3.2 Current Applications in Investment Decisions

The adoption of artificial intelligence in BA investment sectors lagged behind public markets due to early-stage decision reliance on human judgment and scarce structured data about private companies. Furthermore, BAs operate independently and have less access to resources and knowledge to develop and deploy AI. More progress

has been made by VC firms, capable of deploying large teams and expertise, with an added synergy of the long-term benefit of the creation of such systems (Röhm et al., 2022). Furthermore, VCs are likely to be pressured by their investors and stakeholders to innovate and stay ahead of the curve.

Venture capital firms employ AI systems for potential deal sourcing through the analysis of patent documents, news articles, social media data, and startup directories. The systems analyze thousands of opportunities to select ones which fulfill particular investment standards or show characteristics of successful ventures (Röhm et al., 2022; Tewari et al., 2020). AI systems enhance traditional due diligence by evaluating big sets of documents and contracts to spot possible risks or opportunities that human inspectors might overlook during manual examination. These systems use natural language processing capabilities to extract essential terms and detect discrepancies and alert human personnel about potential problems (Tse et al., 2022).

Investors further use machine learning systems to monitor their portfolio companies through combined data analysis from financial reports, customer metrics, employee communications, and market indicators, allowing the detection of warning signals of performance problems. Such systems can also aid with identifying growth possibilities (Trippi & Turban, 1992). The integration of AI technologies within venture capital is expected to evolve significantly (Röhm et al., 2022).

The use of AI in venture exit (sale, IPO, merger) transactions helps investors find the best timing for their exits and determine the most advantageous deal structure by evaluating market trends with comparable deals as well as company-specific elements, thereby maximizing exit value. These applications use diverse data sets together with different modeling techniques to help investors make strategic choices about complex matters (Bahoo et al., 2024; Trippi & Lee, 1995).

2.3.3 Use of AI for Business Angel Decision Making

Some research exists on the use of machine learning (ML) techniques to predict early-stage investor decision-making, but the lack of a rigorous framework (such as the CFA) for analysis limits accuracy. Artificial neural networks (ANNs) and random forests techniques have been employed to predict investment outcomes. Variable such as funding ask, startup valuations, and team composition were evaluated, resulting in a 67% prediction accuracy in yes/no investment scenarios using Shark Tank as a dataset (Deodhar et al., 2024). Logistic regression and neural network models have been used on a similar dataset of Shark Tank deal records, achieving a predictive accuracy of 62.5% in forecasting deal outcomes (Sherk et al., 2019). Furthermore, tree-based models were used to analyze Shark Tank startup valuations and investor decision-making, from show metadata, with an accuracy of 70% (Lavanchy et al., 2022).

Recent developments in explainable AI (XAI), with techniques such as SHapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) have enabled researchers to understand which features and factors are most influential in venture assessments (Barredo Arrieta et al., 2020). Such techniques are particularly important in high-stake environments such as BA investing as decision traceability, intelligence, and interpretability of decisions are important. Recent advanced NLP techniques have been deployed to evaluate various unstructured text data to identify potentially unethical or non-compliant signals to power informed investment decisions (Jagdale & Deshmukh, 2025). Such techniques are especially relevant in highly regulated financial and investment markets.

2.3.4 Ethical Considerations and Bias in AI Investment Systems

AI systems used in investment decisions pose potential ethical concerns regarding fairness as well as transparency, accountability, and systematic risk exposure. The learning algorithms can both sustain and strengthen existing biases found within their

training data sets or model architecture design (Osasona et al., 2024). The investment process could lead to unfair discrimination of certain groups of entrepreneurs, ventures, or business models due to previous data of such groups traditionally receiving unsuccessful investment results. The study of traditional venture capital decision-making reveals multiple forms of bias including gender bias, racial bias, and geographic bias (Greene et al., 2001; Leavy, 2018). AI systems trained on historical investment data may learn and maintain these biases, which could sustain discriminatory capital distribution patterns (Ferrer et al., 2020).

Most AI systems function as black boxes that do not reveal the decision-making processes which support their recommendations. Such shortcomings may create challenges during investment recommendation analysis as stakeholders strive to understand the basis of recommendations. Furthermore, regulatory requirements demand explainable decision-making processes. The implementation of AI systems in investment decision processes creates challenges with assigning responsibility and liability when adverse outcomes occur. Responsible AI implementation in investment requires frameworks which establish human supervision alongside explainability and error management (Dey, 2024; Ehsan et al., 2021).

When investment organizations adopt similar AI systems, they face the risk of creating systemic problems through uniform responses to market signals or identical investor behavior. The early-stage investment process presents high risks, as individual investment choices substantially affect both venture success and entrepreneurial ecosystem development.

2.3.5 Technology Acceptance and Adoption Models

The Technology Acceptance Model (TAM) created by Fred D. Davis (Davis, 1993), further developed by Venkatesh et al. (Venkatesh et al., 2003) in the Unified Theory of Acceptance and Use of Technology (UTAUT), provides a theoretical framework for understanding technology adoption in professional environments.

According to TAM, the adoption of technology is informed by two main factors: perceived usefulness and perceived ease of use. The perceived usefulness outlines the users' belief that the new technology (investors in our case) will enhance their decision quality, efficiency, and consistency. The perceived ease of use assesses how much mental effort the users need to understand and operate the system properly (Davis, 1993). The UTAUT model extends TAM through additional elements such as social influence and facilitating conditions, which also account for moderating factors including experience, gender, and age (Venkatesh et al., 2003).

The adoption of professional investment technology is further informed by peer influence and reputable industry standard practices. Research in recent years shows that trust stands as a vital factor for (AI) system adoption, particularly in high-stakes decision-making environments (Choung et al., 2022; Glikson & Woolley, 2020). The level of trust users have in AI systems depends on their perception of system transparency together with predictability, reliability, and alignment with expected decision-making processes. The adoption rate of AI-augmented decision support systems increases when they use recognized and proven decision-making frameworks instead of creating new and unfamiliar decision-making approaches (Choung et al., 2022; Glikson & Woolley, 2020).

Chapter 3: Theoretical Framework

3.1 Conceptual Model Development

3.1.1 Integrated Human-AI Decision-Making Framework

This study bases its research theory on an integrated system for human-AI collaboration in complicated decision situations. The framework integrates behavioral decision theory and information processing research along with technology acceptance models to demonstrate how artificial intelligence can serve as an enhancement for human expertise during venture evaluation.

Human experts following traditional CFA implementation methods must extract vital information from natural language venture descriptions and evaluate across eight dimensions before combining scores into an overall assessment.

Human experts combine both analytical System 2 processes with intuitive System 1 processes to evaluate ventures. Evaluators utilize their domain knowledge and pattern recognition abilities with explicit analytical frameworks (Kahneman, 2011). The analytical framework remains intact through AI-augmented implementation, though AI augments or replaces human cognitive processes in certain steps (such as grade assignment or information processing).

Large Language Models function as sophisticated tools for natural language processing that extract and understand vital information within unstructured venture descriptions, pitches, or supporting investment data. The integration of multiple evaluation dimensions through the use of machine learning classification algorithms leads to systematic predictions of investment results.

The assessment framework relies on expert-validated knowledge with clear evaluation standards (i.e., CFA), which form its foundation. The system functions as cognitive augmentation by mimicking human information processing and pattern recognition functions through artificial intelligence, without compromising established human decision-making frameworks and practical knowledge.

3.1.2 Information Flow and Processing Architecture

The architecture consists of four primary stages: information extraction, factor assessment, integration and prediction, and validation and feedback.

The source of input for the venture information comes from natural language pitch presentations made by entrepreneurs to investors. An LLM pipeline is used to analyze unstructured information to extract relevant each of the eight CFA factors. At this stage, the system interprets venture details through the lens of evaluation criteria, performing work analogous to human cognitive processing.

For each CFA factor, an independent evaluation is conducted (in isolation from other factors) to extract information matching the CFA evaluation rubric to produce standardized scores. LLM prompting instructs these modules to duplicate the judgment of expert evaluators by applying explicit guidelines and implicit pattern recognition from historical expert evaluation records.

Factor scores are then combined through machine learning classifiers to generate final yes/no investment outcome predictions. Final outputs receive validation through comparisons with available data that includes actual investment results and trained human evaluations.

3.2 Technology Acceptance in Professional Investment Contexts

3.2.1 Adaptation of Technology Acceptance Models

AI-augmented CFA system implementation requires specific modification of established technology acceptance frameworks that address professional investment decision-making contexts. The Technology Acceptance Model (TAM) along with its extensions provide a base for predicting real-world adoption probability. However, adjustment and evolution is required to fit evolving investment decision environments (Davis, 1993). The usefulness of AI for professionals hinges on several key factors, such as better decision accuracy, enhanced efficiency, risk reduction, and competitive market benefits. The adoption of professional investment tools depends on demonstrating actual

value through enhanced decision quality or decreased decision costs rather than consumer-focused convenience or entertainment features (Phillips-Wren, 2012).

Investment professionals show greater concern about making wrong investments (false positive errors) than about overlooking profitable opportunities (false negative errors). AI systems which minimize overall errors while keeping acceptable false positive rates tend to receive better perceived usefulness evaluations than systems that maximize overall accuracy without accounting for error type preferences (Buczynski et al., 2021; Trippi & Turban, 1992).

Professional investment workflows demand that usability goes beyond user interface simplicity, as they require integration capabilities with existing processes or systems to ensure minimal disruption to existing decision systems.

High-stakes professional investment decisions produce major financial impacts as well as potential risk for damage to professional reputations. Technology acceptance at this level requires maximum reliability standards and visible decision-making methods which support human oversight and accountability. The rejection of "black box" AI systems will remain high regardless of their performance results as long as these systems provide unexplained recommendations.

3.2.2 Social Influence and Institutional Factors

Early adopters within professional investment communities create strong social influence, affecting how technology acceptance spreads throughout the group. The VC and BA communities maintain unofficial networks for sharing information about investment results along with decision approaches and tool assessment (Goldenberg et al., 2009; Wetzel, 1983). Professional investors track their peers' achievements and the use of decision methods. The adoption of AI decision support systems will be greater when prominent investors within the community endorse or implement these systems.

Regulatory bodies and fiduciary duties force investment professionals to document their decisions and hold them responsible for investment-related losses or damages. AI systems need to deliver proper documentation alongside transparency

standards to fulfill regulatory standards and safeguard investors from legal responsibility claims (Koshiyama et al., 2021).

3.2.3 Performance Expectations and Validation Requirements

Investment professionals use multiple evaluation benchmarks to assess AI systems which include their individual historical results and peer investor results and theoretical best results. Systems must outperform human decision-making while preserving or improving current practice performance levels (Hernández-Orallo, 2017).

The validation evidence required by investment professionals may include back testing historical data along with expert human evaluator comparisons and performance stability tests across various market conditions and venture types (Hernández-Orallo, 2017). Artificial Intelligence systems need to work within existing risk management systems and minimize the creation of new risk factors. Our proposed system provides explicit decision-making outcomes tied to eight factors, alongside their breakdowns, allowing for human intervention and oversight.

3.3 Information Processing Theory Applications

3.3.1 Cognitive Load and Processing Capacity

Human decision-makers possess restricted working memory abilities, especially when handling vast complex information sources at once. The eight-factor CFA framework helps minimize cognitive load through its structured categories for evaluation of information, yet human evaluators may still experience challenges in distributing attention across all factors while reviewing large quantities of assessments simultaneously and tirelessly, without the loss of performance or consistency. AI systems help reduce cognitive load through their ability to process venture data before sorting it according to CFA factor categories, thereby allowing evaluators to concentrate on judgment integration and higher-level assessment instead of information retrieval and

structuring tasks. Through division of work between human and AI systems, evaluation procedures become more efficient and accurate. AI systems have the ability to enhance attention allocation through systematic processing of complete information while eliminating attentional biases (Lindsay, 2020).

3.3.2 Pattern Recognition and Expertise Development

BAs create complex, often subconscious, mental frameworks of successful and unsuccessful ventures, allowing them to quickly evaluate new opportunities through similarities to their past experiences (Baron, 2006).

Research shows that human experts demonstrate exceptional abilities to detect intricate patterns within complex data that may be difficult to articulate explicitly but holds essential value for accurate assessment. The process of expert pattern recognition remains vulnerable to three primary biases which may cause experts to overemphasize memorable examples while neglecting base rate differences and pattern similarity misinterpretations (Jain et al., 2000).

Modern LLMs demonstrate advanced pattern recognition abilities, able to find complex data relationships in text-based information without needing specific pattern definitions. Such LLMs have the capability to detect refined linguistic signals which indicate venture quality that human evaluators might not detect or properly evaluate (Naveed et al., 2023).

3.4 Opportunities and Limitations of the CFA

While the CFA framework has significantly influenced BA decision-making, subsequent research has explored its applications and limitations. Structured evaluation frameworks, such as those modeled after the CFA, enhance decision-making efficiency in BA syndicates during due diligence (Adomdza, 2004).

The strict application of these frameworks may unintentionally reinforce biases or overlook the nuanced potential of atypical, and often successful, ventures. A systematic review of gender disparities in entrepreneurial equity financing indicated that social biases, both conscious and unconscious, can affect BA decisions, especially during early-stage evaluations (Koziol et al., 2025).

Structured tools such as the CFA help reduce social biases by offering objective benchmarks; however, the effectiveness of these frameworks is contingent upon the degree to which their foundational criteria are devoid of inherent bias. Should the critical factors embody implicit social biases (gender, race, language, etc.), structured decision tools may unintentionally sustain, rather than address, inequities in early-stage funding decisions.

The CFA framework effectively provides structured evaluations of early-stage ventures; however, several inherent constraints limit its scalability, efficiency, and broader applicability. The limitations arise from the framework's reliance on manual human assessment (Maxwell, 2009) and dependence on data quality, resulting in challenges especially evident in the fast-paced, high-volume contemporary entrepreneurial finance (Arundale & Mason, 2024). The process is time-consuming and resource-intensive, with expert-driven CFA assessments typically requiring up to six weeks and costing up to $1,400 per evaluation (CFA Snapshot, 2015), creating substantial obstacles for resource-limited startups and organizations with restricted evaluation capabilities (Gonzales M. et al., 2024).

Entrepreneurs must also make decisions rapidly as markets change and opportunities arise, with studies indicating that moving quickly is a winning strategy in innovative early-stage startups. For such a startup to succeed, rapid iterations and decision making (often referred to as a Lean Startup approach, coined by author Eric Ries) must be paired with the ability to focus on the right issues, while ignoring the wrong ones (Ries, 2011).

Early-stage startups do not possess sufficient data to support their business assertions, especially on market validation, financial forecasts, and operational preparedness (Cassar, 2009; Mason & Harrison, 2003). This problem is exacerbated by information asymmetry, when entrepreneurs may inadvertently exclude essential details or, in certain instances, deliberately withhold information to portray their ventures more favorably (Baker & Ricciardi, 2014; Barbaroux, 2014).

A second significant limitation is the subjective nature of human assessments. Cognitive biases, including anchoring, overconfidence, and availability heuristics, can still impact investment decisions, despite the structured criteria inherent in the CFA. Kahneman emphasized that even skilled evaluators are vulnerable to biases that can impair judgment, especially in high-uncertainty situations that demand swift decision-making (Kahneman, 2011). This challenge is intensified in competitive entrepreneurial ecosystems, where swift evaluations are crucial for seizing time-sensitive opportunities and sustaining deal flow momentum (Kliger & Kudryavtsev, 2010).

3.5 Hypothesis Development

This research was designed to explore the use of trained AI models (leveraging previous work on the CFA) to provide a solution that addressed the cost, training, and timing constraints that contributed to CFA's limited deployment. If successful, this novel approach would allow the deployment of an AI-powered CFA to benefit fund-seeking entrepreneurs and the broader entrepreneurial ecosystem.

The first question to address was whether a trained LLM can replicate the manual CFA assessment in providing valuable insights, previously exclusively done by trained experts. Given the structure of the CFA (i.e. each factor must be independently validated), and the results of our initial analysis, we propose the following hypothesis:

> **H1a:** When evaluating startup pitches, a CFA-augmented AI model would produce consistent and accurate CFA evaluations across each of the eight critical factors when compared to the scoring of trained human evaluators.

The nature of the eight factors in the CFA is that they are non-compensatory; thus, a positive investment decision is only made if there is no fundamental flaw in any one of the factors, allowing the CFA to predict the actual investment decision on this basis. This allows us to propose the following hypothesis:

H1b: When forecasting an investment decision of a group of BAs, the CFA-augmented AI model will accurately predict the outcome.

There is no doubt that commercially available AI LLM models (such as ChatGPT, Gemini, and Claude) have been effective in helping entrepreneurs develop their ventures and improve every aspect of their business and pitch through the insights they provide. It therefore challenges the usefulness of the CFA-augmented LLM, as it could produce similar results and advice to an unprompted LLM. While it is likely that an LLM is inherently less biased than any randomly selected individual, and can produce fast and cheap feedback, the whole basis of the previous research on the validity of the CFA tool is that this structured approach, analyzing the eight critical factors, provides more accurate decision predictors and advice than an unstructured approach. This leads us to our final set of hypotheses:

H2a: A CFA-prompted LLM would produce better evaluations and investment predictions than an unprompted "stock" LLM (i.e., out-of-the-box GPT-4).

H2b: A CFA-prompted LLM would produce better and more relevant feedback and advice to the entrepreneur than an unprompted "stock" LLM.

Chapter 4: Research Methodology

4.1 Research Philosophy and Design

4.1.1 Philosophical Foundations

The study employs a positivist approach as it seeks to find objective knowledge about investment decision-making through systematic empirical research. A positivist paradigm was adopted because the study tests specific hypotheses about AI-augmented assessment effects on investment outcome prediction through quantifiable performance and reliability measures.

Both exploratory and explanatory elements are present in the design. The explanatory section of the research evaluates if AI-augmented CFA assessment achieves predictive accuracy similar to human trained evaluators, while testing performance against established benchmarks. The exploratory component investigates how different CFA factors affect AI-based assessment and what practical effects AI augmentation produces for diverse stakeholder groups.

An underlying assumption is that venture quality and investment potential are measurable, but accepts that systematic evaluation will produce uncertain results with unavoidable measurement errors. The assumption supports using the validated CFA framework for standard comparisons and treating investment outcomes as objective decision quality measures.

Business ventures are treated as entities whose measurable attributes determine their success potential and investor appeal. Venture outcomes depend on factors beyond initial characteristics (market changes, psychology, life events, competition), but we assume that initial assessment of ventures delivers valuable information about their future performance.

4.1.2 Research Design Framework

The research design consists of multiple phases, enabling a thorough assessment of the AI-enhanced CFA pipeline performance throughout various measurement dimensions and validation methods. Controlled experimental approaches were combined with field-study methods to achieve both high internal validity and practical application.

The first phase of System Development and Training selects and refines AI models for CFA factor assessment through prompt engineering alongside traditional ML techniques. The technical base required for future validation tests is created during this phase.

The second phase evaluates how well AI-generated CFA assessments match human expert evaluations through the use of correlation measuring techniques.

The third phase investigates predictive validation through the use of Shark Tank dataset, analyzing how AI-generated assessments perform relative to actual investment choices (yes or no investment outcomes). This phase demonstrates how well the system works in real-world prediction scenarios.

The fourth phase evaluates framework-guided LLM (CFA) assessment against generic LLM (out-of-the-box LLM without a specific framework) evaluation methods when analyzing identical venture information.

Three validation methods were employed: correlation analysis, classification performance metrics, and system output qualitative assessment. Triangulation was used to build robust evidence about system performance despite acknowledging each evaluation method's individual limitations.

4.1.3 Unit of Analysis and Sampling Framework

Our dataset, consisting of 600 pitches from the TV show Shark Tank, functions as an appropriate sampling frame for research as it provides several advantages for analysis. The dataset features standardized presentation formats alongside known investment outcomes. The pitches emulate real authentic investor-entrepreneur (as the Sharks actually invest funds in ventures) interactions that researchers can access for research purposes.

4.2 Data Collection and Preparation

4.2.1 Primary Data Sources

The combination of public Shark Tank data alongside trained human evaluations enables both extensive statistical analysis and performance benchmarking.

The "Shark Tank US Dataset" available through a public Kaggle repository, consisted of 1,153 venture pitch metadata records (of which 600 were used, based on transcript and resource availability) that track ventures from the television program.

The dataset contains episodes ranging from seasons 2009 through 2024, covering industries such as technology, consumer products, food and beverage, health and wellness, and professional service. The pitch transcripts were obtained by transcribing publicly available show episodes. The transcripts include all spoken dialogue (aired on TV) and investor-entrepreneur exchanges, consisting of unstructured text fit for LLM assessment. We ensured to post-process transcripts through the standardization of formatting and encoding, removing stage directions and music cues, and omitting discussions of investment negotiation and decisions.

4.2.2 Trained Human Assessment Collection

To evaluate the performance of our AI pipeline, we obtained CFA pitch evaluations from human evaluators for a set of randomly selected Shark Tank pitches

from our dataset. Three university graduate students received training using CFA literature from the Canadian Innovation Center (CIC), emulating Andrew Maxwell's evaluation criteria and processes. Each evaluator independently evaluated 31 venture pitches through the complete eight-factor CFA framework.

4.2.3 Data Preprocessing and Feature Engineering

The gathered data required extensive data cleaning and feature engineering to become suitable for machine learning model training and analysis. Venture pitch transcripts were preprocessed in three steps: normalizing capitalization and punctuation, followed by commercial break and introductory segment removal and segmentation into analytical units. Preprocessed data underwent systematic consistency checks, outlier detection, and verification against its original sources.

4.3 AI Model Development and Training

4.3.1 Large Language Model Selection and Configuration

We employed multiple state-of-the-art LLMs for AI-augmented CFA assessment, which allowed comparison of different model architectures and capabilities. Models were selected based on their performance benchmarks and API availability.

Four primary LLM variants were selected: GPT-4, GPT-4.1, GPT-4.1-mini, and GPT-o3, representing different generations and optimization methods of the GPT model family. Each model presents a different capability-speed-cost trade-off, which help find the best solution for deployment needs.

Various LLM configurations were tested by adjusting model temperature settings for response variability, token limits for response length, and prompt structure for assessment focus and format. The optimization process involved systematic testing of various input formats and parameter settings.

4.3.2 Prompt Engineering and Framework Implementation

The study employed systematic prompt engineering to instruct LLMs in performing venture assessments through CFA framework criteria. Consistency and accuracy heavily depend on prompt engineering.

Complete prompt templates were developed, containing factor definitions alongside assessment criteria, scoring guidelines, and example evaluation samples for each CFA factor. These prompts give LLMs complete directions to perform evaluations that match human evaluator standards.

Each prompt template includes the following elements:

- Factor Definition: Clear explanation of the factor (based on CFA literature) being assessed and its role in overall venture evaluation
- Assessment Criteria: Specific guidelines for evaluating the pitch according to the CFA literature and rubric
- Scoring Guidelines: Detailed rubric for assigning letter grades from A+ to C- (and N/A)
- Recommendation Guidelines: Instructions for how to provide specific and relevant recommendations for the improvement of the factor
- Output Format: Structured format requirements for assessment responses

The optimization process involved the creation and testing of multiple prompts, structures, and variations, while adjusting assessment criteria based on first-round results and checking prompt performance against human evaluation.

Multiple quality-enhancing design principles emerged from the validation process: explicit definition per grade and -/+ assignment, orchestration of eight factors in separate silos, separation of grade, evaluation, and recommendation prompting, and temperature settings. Function-calling was used to ensure the LLM OpenAI API returned predictable and consistent results per each grading criteria.

Context chunking methods were developed to handle lengthy pitches by dividing information into sections which maintain crucial relationships between sections.

Chain-of-thought reasoning was used to perform methodical step-by-step evaluation and reasoning, which produced higher quality assessments.

4.3.3 Hypothesis 1a: Synthesizing CFA Features Using AI

We optimized multiple versions of GPT models (GPT 4, 4.1-mini, 4.1, o3) to evaluate each pitch transcript against the CFA rubric by employing advanced techniques, such as agent/workflow orchestration, function calling, context chunking, and prompt engineering. The custom model assigned grades according to the CIC's standardized rubric, with scores varying from A+ to C- or N/A (Table 1 and Table 2) (Maxwell et al., 2011). The grades were subsequently codified into structured numerical formats (0 to 10) to facilitate machine learning analysis (Table 2). Maxwell's research indicated that the numbers 3 and 7 were intentionally excluded from the CFA scale to signify that each grade transition denotes a substantial increase in investment readiness, thereby ensuring that scoring differences accurately reflect variations in startup quality (Maxwell et al., 2011).

To achieve human-graded CFA scores, we trained three university students to evaluate and grade startup pitches using the CIC's literature and rubric concerning CFA. The trained students then graded the same 31 text transcripts from our database of 600 Shark Tank pitches across the eight critical factors. For each pitch, we calculated the arithmetic mean of the three graders' CFA scoring to achieve score consensus.

Table 1. Critical Factor Assessment (CFA) Criteria

Factor	Criteria	Evaluation Questions
Features & Benefits	Performance Advantages	Does the product offer competitive advantages over current solutions?
	Benefits & Costs	Are benefits substantial relative to costs?
	Customer Demands	Does it meet specific customer needs effectively?
Readiness	Product Delivery	Is the product ready for market, with key milestones achieved?

	Validation Tests	Are there beta tests or customer validations supporting readiness?
Barriers to Entry	Uniqueness	Does the venture have proprietary tech or strong IP protection?
	Market Differentiation	Are there clear competitive advantages?
Adoption	Customer Engagement	Is there evidence of customer interest or early adoption?
	Market Validation	Have customers committed to purchasing?
Supply Chain	Operational Readiness	Are supply chains and partnerships well-established?
Market Size	Revenue Potential	Is the market size large enough to support high growth?
Entrepreneurial Experience	Industry Expertise	Does the team have relevant industry or startup experience?
Financial Expectations	Financial Viability	Are the projections realistic and sustainable?

Table 2. CFA Grade Conversion Table

Grade	Score	Adjusted Score
A+	10	80
A	9	72
A-	8	64
B+	6	48
B	5	40
B-	4	32
C+	2	16
C	1	8
C-	0	0
N/A	0	0

We analyzed the AI-CFA's ability to evaluate startup pitches by comparing the 31 expert CFA grades (sum of eight factors) to the AI-CFA grades (using four different models: GPT 4, 4.1, 4.1-mini, o3) across three correlation metrics: Spearman's ρ (monotonic rank agreement), Pearson's r (linear correlation), and mean absolute error (average point-wise deviation). This resulted gaining the ability to determine how accurately a CFA-prompted LLM can replicate human evaluation.

4.3.4 Hypothesis 1b: AI Model Assessment of Features

We employed a multi-model ML pipeline to predict the likelihood of securing funding (binary outcome: deal/no deal) based on the synthesized CFA features. Data preparation began with exploratory data analysis to visualize the dataset and gain insights into existing patterns and characteristics. We discovered a class imbalance, with 66% of pitches resulting in a Yes (deal) investment, which we attempted to address through data sampling techniques.

All classification models were trained with a fixed random seed (random_state = 42) to ensure reproducibility. Performance metrics are reported from stratified cross-validation procedures.

Next, we generated model features with different CFA factors (i.e. all 8 factors, 8 choose 7, etc.), as well as the requested funding amount and its corresponding equity conversion (Ask $, Ask %) to test the predictive efficacy of different combinations. While Maxwell's research yielded the highest accuracy with all the eight factors (Maxwell et al., 2011), it was prudent to conduct independent evaluation with our model. We chose the following popular, accessible, and efficient classification algorithms for training and evaluation: Logistic Regression, Support Vector Machines (SVM), Naive Bayes Variants, Decision Trees and Random Forests, Gradient Boosting, AdaBoost, XGBoost, CatBoost, Neural Networks, and Linear Discriminant Analysis (LDA). To optimize model performance, we employed hyperparameter optimization within a cross-validation framework, testing different configuration and settings for each model across smaller splits of data, which help with overcoming potential model overfitting. We also tested various ensemble methods, which are common stacking techniques involving the use of multiple base classification models (E.G Random Forests and Logistic Regression) in combination. Within the cross-validation optimization, top-performing models were then retrained on the full dataset and tested on a holdout test set (10–20%) to assess generalizability and overall predicative accuracy.

Models were evaluated using the following metrics:

1. Accuracy: Of all predicted investment outcome, what proportion was correct.

2. F1 score: Balancing precision and recall, especially important for imbalanced datasets.
3. Specificity: Measuring the ability to correctly identify negative cases (i.e., predicting “no deal” accurately).
4. Balanced score: A custom metric introduced by us averaging F1, accuracy, and specificity for during cross validation, to optimize for a model that balances all metrics.
5. ROC AUC: The probability that the model will assign a higher “funding likelihood” score to a randomly chosen “deal” pitch than to a randomly chosen "no deal” pitch.

4.3.5 Hypothesis 2a/2b: Untrained “Stock” LLM Pitch Processing

We initiated a ChatGPT chat using GPT-4 with the following prompt:

“Attached are transcripts from Shark Tank pitches, where startups are seeking investments. For each pitch, please analyze and provide a recommendation on whether an investor should consider funding. Include specific reasons for your recommendation based on the pitch details. Additionally, offer targeted advice to each entrepreneur on how they could refine their business model or pitch to enhance their chances of securing an investment. Finally, clearly state a “deal” or “no deal” decision on whether an investor should fund this startup.”

The prompt was accompanied with a Shark Tank pitch transcript, which was repeated with 31 pitches. We created a new incognito evaluation session for each pitch to ensure no previous memory or conversations would affect the result. Each evaluation, recommendation, and deal/no-deal decision was recorded for comparison with our CFA-AI model. The deal/no-deal predictions were compared with the known deal/no-deal outcomes, resulting in a confusion matrix comparing the untrained LLM’s predictive capabilities. The text evaluation and recommendation were read and manually compared for length of feedback, contextual relevance, depth of insight, and perceived helpfulness.

Chapter 5: Results

5.1 Hypothesis 1a/1b: CFA-AI feature extraction and deal prediction

Across our two-correlation metrics, the CFA-AI factor-grader powered by GPT-4.1-mini, outperformed the rest, and showed a very strong ability to emulate trained human labelling across the eight CFA factors. Table 3 compares the CFA scores (sum of eight) of four GPT model variants against human consensus on a set of 31 pitches.

Table 3. Comparing CFA-AI and Trained Human Grading

Model	Spearman's ρ	p (Spearman)	Pearson's r	p (Pearson)	Mean Absolute Error (pts)
GPT 4.1-mini	0.909	1.58×10^{-12}	0.892	1.60×10^{-11}	5.9
GPT o3	0.847	1.85×10^{-9}	0.849	1.65×10^{-9}	3.9
GPT 4	0.796	8.46×10^{-8}	0.792	1.13×10^{-7}	9.0
GPT 4.1	0.786	1.59×10^{-7}	0.792	1.10×10^{-7}	8.6

When evaluating monotonic rank agreement (Spearman's ρ), GPT 4.1-mini achieved $\rho = 0.909$, indicating that its ranking of pitches closely mirrored trained evaluator grading (values > 0.8 are considered very strong). Linear correspondence (Pearson's r), $r = 0.892$, showed that the model's raw score predictions lie close to a best-fit line against the trained evaluator scores. For mean absolute error, an average absolute deviation of 5.9 points on an 80-point scale (eight CFA factors x 10 highest possible score) meant GPT 4.1-mini is within ~5 % of the human consensus (derived as a simple arithmetic mean of our three human evaluators' total CFA score).

We deployed bootstrap significance testing (10,000 iterations) to stress test the repeatability of our results. GPT-4.1-mini significantly outperformed GPT-4.1 ($\Delta\rho = 0.122$, $p = 0.045$). The differences between GPT-4.1-mini and GPT-o3 ($\Delta\rho = 0.069$, $p = 0.124$) and GPT-4 ($\Delta\rho = 0.115$, $p = 0.079$) were not statistically significant at $\alpha = 0.05$. The limited holdout sample ($N = 31$) constrains statistical power for detecting moderate correlation differences.

To evaluate human agreement (correlation) per factor across all eight CFA dimensions, Table 3b reports per-factor AI–human correlations for the top-performing model (GPT-4.1-mini). Correlations were strongest for Readiness ($\rho = 0.650$) and Financial Expectations ($\rho = 0.623$), and weakest for Features & Benefits ($\rho = 0.168$) and Supply chain ($\rho = 0.171$). Six of eight individual factors achieved statistical significance ($p < 0.05$). Crucially, per-factor noise averages out in the total CFA score, explaining the substantially higher total-score correlation ($\rho = 0.909$).

Table 3c demonstrates pairwise agreement among the three trained human evaluators on the same 31-pitch holdout set. Interestingly, the mean pairwise human–human agreement on total CFA score ($\rho = 0.465$) is substantially lower than the AI–human correlation ($\rho = 0.909$). This indicates that the CFA-prompted AI model produces assessments that are more consistently aligned with the group consensus than individual human evaluators are with one another. The AI outperforms human–human agreement on six of eight individual factors as well. This finding confirms that the AI model operates well within, in fact exceeding, the variability of trained human assessors.

Table 4. Per-Factor GPT-4.1-mini vs. Human Consensus (N = 31)

Factor	**Spearman ρ**	**p (Spearman)**	**Pearson r**	**p (Pearson)**	**MAE**
Total (CFA Sum)	0.909	1.58×10^{-12}	0.892	1.60×10^{-11}	5.87
F1 – Features & Benefits	0.168	0.367	0.039	0.837	1.39
F2 – Readiness	0.650	7.70×10^{-5}	0.620	1.97×10^{-4}	1.35
F3 – Barrier to Entry	0.618	2.12×10^{-4}	0.594	4.22×10^{-4}	2.24
F4 – Adoption	0.448	0.012	0.387	0.032	1.69
F5 – Supply Chain	0.490	0.005	0.461	0.009	3.39
F6 – Go-to-Market	0.171	0.357	0.146	0.434	2.82
F7 – Entrepreneur Experience	0.493	0.005	0.513	0.003	2.15

F8 – Financial Expectation	0.623	1.82×10^{-4}	0.747	1.40×10^{-6}	2.0

Table 5. Human–Human vs. AI–Human Agreement (Spearman ρ), N = 31

Metric	**Mean Pairwise ρ (Human–Human)**	**AI–Human ρ (GPT-4.1-mini)**
Total (CFA Sum)	0.465	0.909
F1 – Features & Benefits	0.091	0.168
F2 – Readiness	0.305	0.650
F3 – Barrier to Entry	0.453	0.618
F4 – Adoption	0.361	0.448
F5 – Supply Chain	0.396	0.490
F6 – Market Size	0.480	0.171
F7– Financial Expectations	0.593	0.493
F8 – Pitch Presentation	0.113	0.623

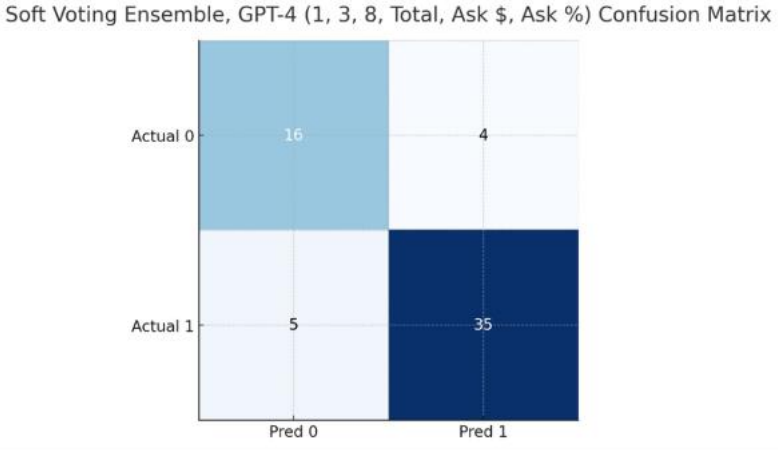
Of the classification algorithms evaluated to test H1b, Model 1 - Soft Voting Ensemble, consisting of Naive Bayes, Logistic Regression, Random Forest, and Gradient Boosting with GPT-4, consistently demonstrated superior performance while utilizing Features & Benefits, Barrier to Entry, Financial Expectations, the total sum of all factors, Ask $ amount, and the respective company equity investors would get. The primary performance metrics for the leading model demonstrated an impressive accuracy of 85.0%, F1 Score of 0.83, Specificity of 0.80, ROC AUC of 0.82, Precision of 0.83, and recall of 0.84 (Table 4, Figure 1 for confusion matrix). Additional models also showed noteworthy performance, as seen in Table 4. Model 2 corresponds to CatBoost (GPT-4.1-mini).

Table 6. Performance Metrics for Top Models

Model	**Feature Combination**	**Accuracy (%)**	**F1 Score**	**Specificity**	**ROC AUC**	**Precision**	**Recall**
GPT-4 (Ensemble)	1, 3, 8, Total, Ask $, Ask %	85.0	0.83	0.80	0.82	0.83	0.84
GPT-4.1-mini (Catboost)	1, 4, Total, Ask $, Ask %	85.0	0.82	0.70	0.76	0.84	0.81

GPT-4.1-mini (Catboost)	All 8, Total, Ask $, Ask %	78.0	0.75	0.6	0.74	0.76	0.74
GPT4 (Ensemble)	All 8, Total, Ask $, Ask %	75.0	0.72	0.65	0.76	0.72	0.73

Figure 1. Soft Voting Ensemble (GPT-4) Confusion Matrix



A feature importance analysis on our top performing stacking ensemble model showed that the total CFA score (importance weight = 0.195) was the single strongest predictor, followed by Financial Expectations (0.158), Ask Amount (0.148), Features & Benefits (0.145), Barrier to Entry (0.101), and Ask Equity (0.077). The results demonstrated that the model effectively identified both successful (deal) and unsuccessful (no deal) pitches with 85.0% accuracy, exhibiting strong performance across all key ML metrics, and supporting H1a/H1b directly. If the LLM model was not effectively assigning CFA grades, the overall predictive accuracy would have suffered.

5.2 Hypothesis 2a/2b: CFA-Prompted LLM vs. Unprompted LLM

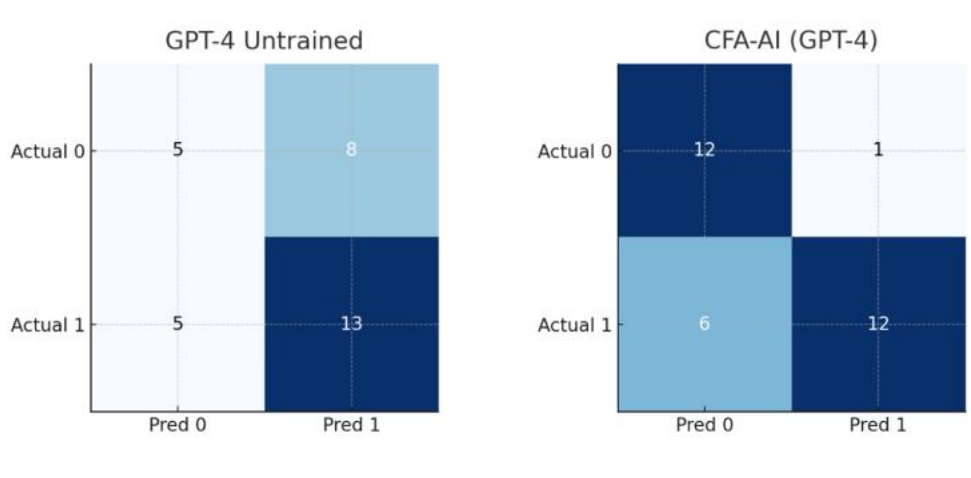
While an untrained GPT-4 showed some predictive capability (accuracy of 58.1%), it was outperformed by our CFA-prompted LLM on the same test dataset (N=31, Accuracy = 77.4%) (Table 5 and Figure 2). The largest predictive gap was seen in the Specificity (of all Shark Tank “no deal” pitches, the percentage our model correctly

labeled as “no deal”) of the two models (0.39 vs. 0.92). Furthermore, CFA-LLM generated contextual evaluations and recommendations in full sentences and paragraphs, addressing various aspects of the pitch in detail, drawing specific examples from the pitch and generating relevant and sophisticated recommendations. In comparison, the GPT-4-generated evaluations and recommendations were presented in short bullet-point format, and were not consistent across various reports, often omitting crucial feedback or insights. The results therefore support H2a/2b.

Table 7. Prompted vs. Unprompted LLM Performance (N=31)

Metric	GPT-4 Unprompted	CFA-Prompted (GPT-4)
Accuracy	58.1%	77.4%
Precision (Deal)	0.62	0.92
Recall (Deal)	0.72	0.67
F1 Score	0.55	0.77
Specificity (No Deal)	0.39	0.92

Figure 2. CFA-Prompted vs. Unprompted Confusion Matrices (N=31)



5.3 Error Analysis

For the 31 pitches and 8 factors evaluated, Factors 1, 3, 5 exhibited the largest mean difference when comparing to human evaluations (Table 6). For the same 31 pitches, Table 7 displays the seven for which the final deal outcome was misclassified (FN = False Negative, FP = False Positive).

A positive difference indicates the AI assigns a higher (more generous) score than the trained human consensus. The consistently positive differences for F5 (Supply Chain), F3 (Barrier to Entry), and F1 (Features & Benefits) suggest that the AI tends to score these factors more optimistically than human evaluators.

Table 8. Largest CFA Differences, CFA-AI vs. Humans

Factor	Mean difference (pts)	Pearson *r*
F5 – Supply Chain	+3.2	0.49
F3 – Barrier to Entry	+2.5	0.43
F1 – Features & Benefits	+1.6	0.38

Table 9. Misclassified Deal Outcomes CFA-AI

Pitch #	Outcome	AI Confidence (> 50 %=Deal)	GPT-4 total	Human total	Error type
3	1	35.8 %	57	57	FN
12	1	17.1 %	27	41	FN
14	1	43.2 %	51	50	FN
17	1	48.3 %	44	42	FN
27	1	20.5 %	36	40	FN
29	1	18.3 %	37	45	FN
10	0	50.5 %	61	51	FP

Chapter 6: Discussion

6.1 Theoretical Implications and Contributions

6.1.1 Advancing Human-AI Collaboration Theory

The AI-augmented CFA results show that AI can closely replicate trained humans when using a validated decision-making framework to evaluate BA decision making, while offering a solution to scalability and consistency issues that limit human-only methods.

The research supports cognitive augmentation theories, demonstrating that structured frameworks built by human experts outperform (out-of-the-box) LLM AI independent evaluation criteria. The results indicate that framework-guided LLMs (77.4% accuracy) surpass generic LLM analysis (58.1% accuracy) on a test size of 31, which validates our hybrid approach. Our research suggests that future AI implementation preserves the theoretical knowledge and practical expertise found in established expert practices.

The research suggests that AI systems can reduce both human cognitive distortions and specific AI systematic flaws. AI systems inherently perform CFA criteria evaluation with consistency, thus eliminating human biases such as mood effects, availability heuristics, and anchoring but produced potential new system errors based on training data patterns and context interpretation restrictions.

6.1.2 Information Processing Theory Extensions

The research outcomes serve as empirical proof for how information processing theory functions in investment decision-making scenarios. The AI system proves capable of extracting important information from venture pitches and structuring it through defined evaluation criteria, thus showing AI's ability to overcome human information processing deficiencies while keeping validated knowledge frameworks intact. As an

effect, humans can dedicate their attention to strategic initiatives and follow-on due diligence.

6.1.3 Technology Acceptance Theory Validation

Research results offer useful knowledge for technology acceptance theory in professional investment settings, but complete validation demands further studies over longer periods of time.

The advantages of AI-augmented CFA (predictive accuracy, efficiency, cognitive load reduction) demonstrate perceived usefulness that should drive technology adoption as predicted by TAM. However, several barriers to perceived usefulness exist, including investors' concerns about transparency, accountability, and performance stability across evolving market conditions. The systematic biases that may be present in AI assessment systems (due to intrinsic biases present in the data set or during reinforcement learning) decrease the perceived usefulness of investors who demand transparent decision-making processes.

The successful use of AI-augmented CFA using natural language processing shows high potential to support perceived ease of use since the system requires little to no technical expertise and can work with existing evaluation systems.

6.2 Methodological Contributions and Innovations

6.2.1 Novel Approaches to AI-Human Collaboration

Our research proposes an approach that can be replicable to additional domains beyond venture assessment. The approach includes three essential elements for framework implementation: clear criterion specification, explicit scoring rubrics, and structured output formats which help with analysis and validation. The research proposes solutions to major LLM deployment obstacles through evaluation consistency between multiple assessments, while offering solutions for handling boundary conditions and

integration with established analytical systems. The methods established here can be transformed into different assessment situations including legal, education, health, professional services, and more. The research also showcases successful methods for uniting various LLM models, human frameworks, and traditional ML to achieve top performance in multiple assessment criteria; Systematic combination strategies help the ensemble approach utilize the strengths of different models by reducing their individual weaknesses.

6.2.2 Bridging Academic Research and Practical Application

The research shows successful integration of academic frameworks into functional applications which resolve practical issues in real-world deployments while maintaining their theoretical concepts; Our CFA-AI system preserves the theoretical structure and empirical validation of CFA while addressing scalability problems affecting its widespread adoption. While this research focused on the CFA, other academic frameworks and methodologies may be deployed. The research proves how sophisticated analytical tools with advanced AI capabilities make them available to a wider audience beyond those who were restricted by resource and expertise limitations. The democratization approach shows promise for other fields that face geographical or economic barriers to domain knowledge and assessment capabilities.

6.3 Comparison with Andrew Maxwell's CFA

This research's core methodology, combining AI with the CFA, allowed us to build on Maxwell's research on BA decision making and propose a novel, more efficient, scalable, and accessible approach to assessing early-stage ventures. To evaluate AI's validity to deploy Maxwell's CFA, we first determined whether it could replicate human-CFA-trained results. Maxwell's CFA method, dependent on human evaluators, demonstrated notable predictive performance, attaining an overall accuracy of 87.5% in forecasting early-stage investment decisions (Maxwell et al., 2011). Our integrated CFA-

AI model demonstrated a test accuracy of up to 85% when employing ensemble voting models to predict BA funding outcomes.

It is important to highlight several differences in data sources and evaluation methods: Maxwell's study reviewed extensive conversations between entrepreneurs and investors, lasting 30–60 min, offering evaluators more context and potential signals for assessment (Maxwell et al., 2011). By contrast, our model utilized a dataset of brief 3–5-min Shark Tank pitches, which offer significant reliability (as discussed earlier in the research) but could still lack depth (via omission of important in-person interactions and knowledge sharing for the purpose of TV entertainment). Maxwell's model combined evaluations from several trained CFA raters, adding a dimension of human consensus that likely enhanced its accuracy (Maxwell et al., 2011), which can be mimicked in our model through the future integration of multi-agent models. Our CFA-AI model utilized a single output generated by an LLM for each pitch, which may introduce variance stemming from model-specific biases.

Maxwell's model demonstrated efficacy in predicting negative decision outcomes, achieving 100% accuracy in identifying pitches that were likely to be rejected based on fatal flaws (compared to our model's 80% specificity) (Maxwell et al., 2011). The CFA-AI model demonstrated enhanced efficacy in forecasting favorable results (Yes Deal). This divergence may suggest variations in the data context. Pitches aired on TV often highlight success-oriented narratives designed to attract investor interest and audience engagement. Maxwell's CFA study employed a dataset of pitches that were not solely broadcasted, thus providing a more precise depiction of the deal/no deal distribution.

6.4 Comparison to Other Shark Tank Studies

Our model's performance surpassed that of previous studies utilizing Shark Tank data, with predictive accuracies ranging between 62.5–70%:

- Deodhar et al. utilized ANNs and random forests on a dataset of 117 pitches from Shark Tank India, achieving an accuracy of 67% (Deodhar et al., 2024).

- Sherk et al. applied logistic regression and neural networks to Shark Tank data, achieving a predictive accuracy of 62.5% (Sherk et al., 2019).
- Lavanchy et al. employed decision tree models to analyze startup valuations and investor decision-making, attaining an accuracy of 70% (Lavanchy et al., 2022).

The improved performance of our model can be attributed to two main factors. First, while prior research frequently relied on structured metadata, such as deal size, team demographics, and industry type, our methodology utilized LLM-based feature extraction to examine the entirety of the textual content in startup pitches. This method captures nuanced and context-dependent information that structured metadata may overlook. Second, by integrating the CFA framework as a systematic evaluative tool, our model focuses on empirically validated, domain-specific factors instead of generic indicators. The integration of comprehensive, text-based features with the structured CFA framework represents a notable methodological improvement, enhancing our work's predictive accuracy relative to previous methods.

6.5 Analyzing the Results

6.5.1 Top Performing Model

The top performing model, combining the LLM-based CFA factor synthesis with soft voting ensemble, exhibited a Recall and F1 score of 0.84 and 0.83 respectively. Recall accounts for the total pitches that received an investment, compared to how many our model identifies correctly (TP/(TP+FN)). F1 score denotes a harmonic mean between precision (of pitches that our model thought will receive investment, how many did) and recall. Our model's specificity (TN/(TN+FP)) of 0.80 denoted its ability to correctly identify a significant portion of non-deal pitches. The overall accuracy of 0.85 (85%) denotes the total number of correct predictions divided by total number of pitches in test set. In a test set of 60, the model thereby correctly classified 51 pitch outcomes. The results indicate that the model did not blindly classify all pitches as "deal" cases, which is often a sign of an overfitted models, but effectively differentiated between investment outcomes.

A potential explanation for the strong performance of Naive Bayes (as a top performing base model in our ensemble) is its robustness with smaller datasets, and the features derived from text. The probabilistic nature enables effective management of class imbalance, and the assumption of feature independence corresponds with the structure of CFA-based evaluation, wherein each factor is evaluated individually prior to aggregation.

6.5.2 Importance of Specific CFA Factors

An analysis of features among the top two performing models identified Features & Benefits (1), Barrier to Entry (3), Adoption (4), and Financial Expectations (8) as the most predictive individual factors, alongside Total CFA Score, Ask Amount, and the respective Ask Percent (e.g. $100,000 in exchange for 10% of the company). Readiness (2), Supply Chain (5), Market Size (6), and Entrepreneurial Experience (7) were missing, which may be explained either by our model's shortcomings, or by the nature of the information typically expressed in the transcript of the Shark Tank interaction. Startups that make it to the show already exhibited some level of Readiness and thereby Supply Chain. Most of the startups displayed strong product readiness and some level of sales, which implies that the product is in the later stages of readiness and basic supply chain processes were in place.

While Maxwell's CFA deemed eight factors important, it is possible that the show producers chose to omit specific discussion points from aired pitches. Certain factors may also have been pre-screened prior to selection of pitches that will air, with Entrepreneur Experience a likely candidate, to ensure the presenter is engaging and relevant. Such factors would limit our model's ability to derive crucial decision-making data, which explains the omission of certain factors from top performing models.

Finally, the training and prompting of our LLM-based feature extraction may have been inadequate for underperforming factors, indicating the need for further finetuning and prompt engineering.

6.5.3 Comparing Performance of Different GPT Models

While GPT-4.1-mini outperformed all other models in its ability to replicate human grading (Spearman's $\rho = 0.909$ vs. 0.796 for GPT-4), it did not replicate its superiority in predicting pitch funding outcomes. Our top GPT-4 based ensemble pipeline outperformed in two key metrics: Specificity and ROC AUC, by 10 and 6 percentage points respectively. This result warrants explanation, as it challenges the common assumption that larger, more capable models (GPT-4 > GPT-4.1-mini) necessarily produce better task-specific performance. Three factors likely account for this result:

First, the CFA evaluation task is rubric-constrained: the model must assign a score on a defined scale for each of eight specific factors, based on explicit criteria. This structured, bounded task rewards precision and adherence to instructions over the broader reasoning capabilities that larger models are optimized for. Second, smaller models tend to produce more concise, focused outputs with fewer extraneous rationalizations. In a rubric-scoring context, verbose reasoning can introduce noise, the model may "talk itself into" a different score by over-weighing tangential considerations. Third, and most importantly, prompt structure and CFA constraints dominate raw model scale as the primary driver of performance. The CFA rubric provides the model with detailed factor definitions, scoring guidelines, and output format requirements, effectively constraining the solution space. Under these conditions, the marginal benefit of additional model parameters diminishes.

As reported in Section 5.1, bootstrap significance testing confirmed that GPT-4.1-mini significantly outperformed GPT-4.1 ($\Delta\rho = 0.122$, $p = 0.045$), while differences versus GPT-o3 and GPT-4 were not statistically significant at $\alpha = 0.05$.

This finding reinforces a central argument of this thesis: it is the validated decision-making framework, not the scale of the underlying model, that drives assessment quality. Framework-guided heuristic evaluation, rather than "bigger is always better," is the mechanism through which AI achieves strong alignment with trained human evaluator judgment.

6.5.4 CFA-Prompted LLM Model vs. Unprompted GPT-4

Our results demonstrated the superior performance and usefulness of a custom-prompted LLM with a research-backed decision-making framework. During initial phases of our research, a common criticism was: “Why is your tool better than simply using ChatGPT?”. This was answered while comparing approaches with 31 pitches (77.4% Accuracy for CFA-AI vs. 58.1% for an untrained GPT-4). While an untrained GPT-4 may excel in a broader context of idea and text generation, it may also suffer from a lack of focused training in identifying investment opportunities correctly, as well as providing optimally useful pitch feedback and improvement recommendations.

6.6 Addressing Cognitive Biases and Decision-Making Enhancement

6.6.1 Bias Mitigation Through Structured AI Assessment

This research suggests that framework-guided AI systems can reduce some sources of human decision inconsistency and cognitive bias by enforcing consistent evaluation criteria; however, it may still inherit data, label, or framework biases from the underlying context (data, framework, prompting, reinforcement learning during training, model architecture, etc.). Our system applies a human-designed framework (CFA) consistently, reducing variability arising from individual human judgment, such as availability bias or anchoring effects, but it does not eliminate all forms of bias, as the framework itself is human-defined and created evaluation criteria. The evaluation process of BAs is vulnerable to mood state and time-dependent factors which impact judgment quality and introduce emotional bias. Our system removes such unpredictable variables, which ensures uniform evaluation quality regardless of external influences that could affect human performance. Organizations greatly benefit from this consistency when faced with a high demand of information processing and deadline turnaround. Structured evaluation criteria may reduce over-reliance on presentation effects; however, demographic fairness was not directly tested in this study due to the absence of labeled demographic data in the dataset. AI systems develop their biases from training data

patterns, with framework-based assessments increasing transparency relative to unstructured judgment by producing factor-level scores and rationales. Future research and deployment must stay alert to algorithmic biases and should also perform regular audits and system calibrations to stop discriminatory practices against particular entrepreneur groups and venture categories.

6.6.2 Enhanced Decision Documentation and Learning

The structured and scalable CFA-AI system lends way to better decision documentation, learning capabilities, organizational learning, and continuous improvement benefits. The system produces complete evaluation documentation for all eight CFA factors alongside detailed records that support accountability needs and regulatory compliance. This system resolves a significant problem in traditional investment evaluation because it ensures that assessment logic gets properly documented for every opportunity. AI assessment systems enable organizations to monitor evaluation patterns throughout time, enabling stakeholders to identify biases and errors while allowing adjustments to assessment criteria through outcome feedback. Our system design further enables users to create standardized evaluation metrics which facilitates meaningful comparison between different ventures, time periods, and evaluators. The benchmarking capability enables users to perform portfolio analysis and performance attribution while supporting strategic planning.

Chapter 7: Limitations and Future Research

7.1 Methodological Limitations

7.1.1 Data Source and Generalizability Constraints

Standardization, access, and scale served as our primary factors for selecting Shark Tank as our research data. Still, several limitations may reduce how well our findings generalize to broader angel investment contexts.

Shark Tank shows are produced for entertainment, which may introduce systematic differences relative to BA evaluation settings. Entrepreneurs pitching on television may adjust their presentations for the audience and entertainment factors, changing the type of information available for AI assessment relative to closed-doors BA interactions.

Television pitches are limited to roughly 3-5 minutes, far less than the multi-session due diligence typical of serious angel deals, limiting our AI models to operate with less information than a real investor would have. This information constraint may understate AI performance in richer-data settings, or alternatively, the structured TV format may have made extraction easier than it would be with unstructured pitch meetings.

Furthermore, television producers select ventures for both investment interest and entertainment value, which likely excludes venture types and founder demographics that are common in broader angel markets. Consumer-facing products are overrepresented as they tend to have clear and relatable value propositions suited to broad television audiences. Complex B2B technologies, financial services, and other ventures that angels commonly fund but that make for less compelling television are underrepresented. AI performance patterns observed here may not hold in broader investment settings as a result.

Geographically, the study covers only U.S.-based ventures and investors. Entrepreneurial and cultural norms, investor expectations, and regulatory frameworks

differ across markets, and we cannot claim these results would generalize internationally without further testing. The exclusive use of English-language data poses another potential constraint; non-English markets would require separate testing and validation.

7.1.2 Temporal and Market Context Limitations

Although our data spans multiple Shark Tanks seasons, the design does not fully account for how shifting market conditions and evolving investor preferences may affect model stability and generalizability. Positively, the dataset covers growth, recession, and recovery periods in the economy, each of which shapes investor risk appetite, valuation benchmarks, and evaluation criteria differently. An AI system trained on historical patterns may not produce reliable evaluations under market conditions it has not encountered.

Emerging technology ventures are particularly challenging here. For sectors like blockchain and AI itself, historical precedents are scarce and evaluation criteria shift rapidly. Angel investor preferences also evolve with market experience, regulatory changes, and industry developments.

7.2 Technical and AI-Specific Limitations

7.2.1 Large Language Model Constraints

The training data used to develop LLMs contains historical text (for example, GPT-4's training cutoff date is September 2021) which may not match modern market conditions, evaluation practices, or venture characteristics, although these challenges are frequently addressed with newly updated models and retraining. The LLM training data contains inherent familiarity biases, potentially favoring ventures that appear in widely available text sources.

When assessments demand specialized domain expertise or technical understanding which is not prevalent in training data, LLMs may produce believable yet factually incorrect information. The assessment errors produced by "hallucination" effects become challenging to spot without expert validation.

Furthermore, the resulting outputs of LLM highly depend on how the input prompts are designed. While the research performed systematic prompt optimization and experimentation, it is nonetheless biased and limited to our version and expertise; Different prompt designs could produce distinct performance patterns which would impact the reliability and reproducibility of the findings.

7.2.2 Data Exposure and Contamination Risk

Shark Tank episodes, transcripts, and deal summaries are publicly available via streaming platforms (e.g., Hulu, ABC), YouTube clips, and fan-maintained databases (e.g., Shark Tank Wiki). It is therefore plausible that the foundation models used in this research were exposed to Shark Tank content during pretraining. However, this research mitigates the risk of simple memorization in several important ways. The evaluation task does not require factual recall of deal outcomes; rather, it requires structured extraction and rubric-based scoring across eight distinct CFA factors, producing factor-specific scores and detailed investment recommendations. The complexity and specificity of this multi-dimensional assessment task substantially reduces the relevance of any memorized episode-level information.

Furthermore, it is important to distinguish this concern from "overfitting" in the traditional machine-learning sense. The base LLMs were not fine-tuned by the researcher on the study dataset; their weights were fixed prior to evaluation. The concern is therefore best characterized as prior data exposure or contamination risk inherent to all foundation-model research using publicly available data.

To fully isolate framework-guided reasoning from potential pretraining contamination, future work should evaluate model performance on non-televised pitch datasets or proprietary angel screening data where no public record of outcomes exists.

7.2.3 Model Interpretability and Explainability Challenges

While the structured evaluation framework offers greater transparency and explainability than out-of-the-box LLM, our approach does not completely reveal certain AI system decision-making processes and evaluations. While we can derive feature

importance for some classification models, such as Naive Bayes, other models' explanations remain unclear. While modern LLMs offer explainability through techniques such as retrieval augmented generation (RAG), referencing specific lines of text in documents, our LLM outputs during feature extraction were not capable of this functionality.

7.2.4 Clarifying the Use of "Bias" in This Research

The term bias is used in multiple, conceptually distinct ways in this thesis. To avoid ambiguity, the following definitions clarify how bias is understood and discussed in this work.

Cognitive (Decision) Bias refers to systematic deviations in human judgment arising from heuristics such as anchoring, availability, representativeness, and confirmation bias. In early-stage investment contexts, such biases are exacerbated by time pressure and limited information, which our CFA-AI system aims to reduce.

Data or Sample Bias arises when the dataset used for analysis is not representative of the broader population of data the deployed system will see. In this research, Shark Tank pitches represent a curated, televised subset of ventures and may over-represent certain industries, presentation styles, or founder archetypes, affecting primarily generalizability rather than internal validity.

Algorithmic Bias describes systematic differences in model outputs that disadvantage certain groups, often due to biased training data or modeling choices. This thesis discusses algorithmic bias as a potential risk. Due to the absence of demographic labels, subgroup fairness was not empirically tested and is therefore identified as an important direction for future research.

Framework (Criterion) Bias arises when evaluation criteria encode normative assumptions about what constitutes "quality" or "investment readiness." While structured frameworks like CFA reduce cognitive noise and inconsistency, they may still privilege certain venture types or business models. This trade-off is acknowledged explicitly.

7.3 Future Research

7.3.1 CFA Factor Synthesis Using OpenAI Competitors

A key contribution of this research is validating the ability of an LLM to extract CFA features from startup pitches. While publicly available LLMs such as GPT-4 are designed with the goal of facilitating broad natural language generation, our research demonstrated that a prompted model could execute complex and specific evaluative tasks. This was made possible through prompt engineering using the CFA rubric, as well as the intelligent architecture of OpenAI API calls, using varying context windows, model temperatures, and so on. Utilizing competitor models, such as Anthropic's Claude or Google's Gemini, may yield different results, as they are built using differing architectures and trained with different datasets. Comparing our GPT-4 and GPT-4.1-mini to additional OpenAI models (o3, o1, etc.) would offer new data and results.

7.3.2 Application for Entrepreneurs and Educators

While not evaluated in-depth in the Results section, our CFA-AI model (deployed as a web application at expitch.com to demonstrate implementability) provided text evaluation based on the criteria for each of the eight factors. It also provided text recommendations for attaining an A+ rating in each critical factor. This resource functions as an educational tool for entrepreneurship courses, which we have validated through dozens of conversations with academics, incubators, and accelerators. Expitch has been used in multiple hackathons by hundreds of students. Future research can examine the usefulness and effectiveness of the CFA in facilitating conversations with startup mentors, by way of providing actionable metrics to track and work on over time.

Future research may include longitudinal studies to evaluate the effects of AI-generated feedback on entrepreneurial outcomes. Controlled experiments comparing groups with and without AI feedback could yield causal evidence regarding the tool's effectiveness. Complementary qualitative research such as surveys can investigate entrepreneurs' perceptions on the tool's helpfulness.

7.3.3 Application for Investors, Policymakers, and Ecosystem Builders

Although seasoned investors often depend on their intuition and judgment for final decisions, the model can function as an efficient triage tool for angel groups, venture capital firms, and accelerators. The deployment of CFA-AI could aid in minimizing cognitive biases, optimizing due diligence processes, and ensuring that valuable opportunities are not missed due to human error or information overload. Although BAs provided valuable inputs during the development of our model, future research could investigate the practical benefits of deploying the CFA-AI tool in real-world investment settings. Rather than solely predicting startup success or failure, this tool has the potential to function as an advanced decision-support system, enabling BAs to efficiently identify promising investment opportunities. Given that human judgment is frequently compromised by cognitive biases such as overconfidence, anchoring, and confirmation bias (Kahneman, 2011), the integration of AI assessments may serve to mitigate these limitations.

To rigorously test this proposition, future studies could employ experimental designs in which angel investors make investment decisions both with and without the assistance of the CFA-AI tool. Additionally, researchers could investigate the impact on deploying AI decision making tools to mitigate intrinsic human bias and subconscious decision-making factors. The insights we provide can guide policy development in entrepreneurial finance, especially in creating programs to support early-stage ventures. The AI system may be utilized in public funding initiatives, startup competitions, or grant evaluations to facilitate objective, data-driven decision-making.

7.3.4 Refinement of Feature Extraction Techniques

Future research should explore additional advanced NLP methods which could further improve predictive capabilities. Sentiment analysis, currently not covered by our GPT-4-based model, can be employed to assess the emotional tone and confidence levels in startup pitches. Techniques such as topic modeling can identify prevailing themes and

subjects in pitches, assessing their alignment with current market trends and investor interests. While we lightly experimented with topic modeling and sentiment analysis, our dataset of 600 pitches was insufficient to capture text patterns.

Future research can investigate the effect of alternative LLMs on model performance using LLMs such as Google's Gemini, Anthropic's Claude, or even SLMs such as ollama qwen, may offer unique architectural differences which may prove to be more performant, such as better contextual comprehension and improved finetuning abilities. More advanced methodologies may include agent-orchestration of multiple different LLMs to facilitate multi-turn discussions of the CFA rubric before assigning a final grade, similar to how the BAs on Shark Tank or Dragons' Den interact. Furthermore, future research can test LLMs with web search enabled, which may offer advantages in the form of real-time web research to validate claims.

7.3.5 Ethical Considerations and Bias Mitigation

As AI models continue to be deployed and used in critical decisions in entrepreneurial finance, future research must thoroughly examine the ethical challenges involved. It is essential to examine algorithmic bias, particularly to assess whether AI models unintentionally introduce or perpetuate biases associated with gender, race, or socioeconomic status in the evaluation of startups. Enhancing transparency and explainability is essential, and both entrepreneurs and investors must be able to interpret AI-generated assessments to foster trust and accountability. Systematic bias audits and the development of comprehensive frameworks for explainable AI can mitigate risks, ensuring that the CFA-AI model is both effective and ethically responsible.

7.3.6 Practical Spillovers

The availability of AI tools to train and educate entrepreneurs will transform how we support entrepreneurship, providing real time on-demand tools that:

- Lower the cost significantly, while improving the speed of feedback. Based on Maxwell's original assessment, the original CFA cost was 3 raters × 2 days or about $1,400 per venture; while the AI version costs < US$1 and is completed in seconds.
- Allow for real-time iteration, providing scores and direct feedback/diagnostics for founders to iteratively learn each component of the CFA. Expitch (our online deployed tool) is already being used in classroom pilots and at industry pitch events.
    - Enhance the use of toolbox and frameworks that have been developed by researchers, enabling a similar prompt-engineering templates to be used to bring other under-used rubrics (e.g., TRL, JTBD, BMC) into daily practice

Chapter 8: Conclusion

This study demonstrated the ability of framework-prompted artificial intelligence systems to support early-stage investment decision making while solving important scalability and accessibility issues. We validated our hypotheses through the creation of a CFA-prompted LLM system, deployed to evaluate and assign Critical Factor Assessment scores (N/A to A+) using a dataset of 600 Shark Tank entrepreneur-investor pitches. CFA-AI scores were then compared to human evaluators trained on CFA literature to validate correlation. Lastly, an ML pipeline was trained, fine-tuned, and evaluated in its ability to predict yes/no investment decision making.

8.1 Summary of Key Findings

The best-performing model achieved 85% accuracy in predicting BA investment outcomes and demonstrated high correlation with trained human evaluators (Spearman's $\rho = 0.909$, $p < 0.001$) across all eight critical factors. We further demonstrated that a CFA-prompted LLM outperformed an unprompted LLM, achieving an accuracy of 77.4% compared to 58.1% respectively. These results confirm that AI performed best when it supports, rather than replaces, human experience within a validated framework.

One of the most significant contribution relates to cost, making AI expert-level assessments widely available and reducing the cost from $1,400 to less than $1, while shortening the review period from weeks to near real-time. We deployed our CFA-AI system as a publicly available web application (expitch.com) and demonstrated feasibility of implementation and adoption by hundreds of entrepreneurs in universities and accelerators.

8.2 Theoretical and Methodological Contributions

This study offers several theoretical contributions. Contributing to the field of behavioral decision theory, our study shows that structured AI systems can reduce cognitive biases when working within a validated framework. Impacting the field of

information processing, our results suggest that AI systems can aid human cognitive overload while maintaining quality of decision making during systematic evaluation (deployed AI systems do not need to sleep or eat, and do not get affected by mood or external events).

The implemented methodology allows complex multidimensional assessment structures to be reproduced using LLMs. We provide replicable methods for rapid engineering and ensemble modeling, as well as validation procedures that can be modified to support other disciplines that require expert-level assessment capabilities. By merging theory and implementation, we show that a validated framework can become a scalable tool without compromising the underlying theoretical principles.

8.3 Practical Implications for Stakeholders

With the deployment and democratization of our CFA-AI tool, expert-level early-stage pitch assessments are now made accessible, providing entrepreneurs around the world with structured feedback on their ventures, no longer constrained by geographic and economic restrictions. The near real-time capabilities of our system enables entrepreneurs to quickly iterate and improve their ventures through structured feedback loops.

The value proposition for angel investors and venture capital firms is efficiency, productivity, scalability, and consistent quality. AI-CFA acts as a high-speed screening layer, saving time in reviewing early deals without impacting quality over large volumes.

Educational institutions and business organizations can use AI-CFA to develop structured educational programs that educate students and young entrepreneurs about investment decisions.

Policymakers and ecosystem developers can use similar tools to develop programs to support entrepreneurs in regions and populations that do not have access to traditional investor networks. Standardized assessment methods support evidence-based policy development and program evaluation.

References

Kane, T. (2010). The Importance of Startups in Job Creation and Job Destruction. Social Science Research Network. https://doi.org/10.2139/SSRN.1646934

Nostrand, E. V. (2025). Small Business and Entrepreneurship in the Post-COVID Expansion. https://home.treasury.gov/news/featured-stories/small-business-and-entrepreneurship-in-the-post-covid-expansion

Kritikos, A. S. (2015). Entrepreneurship and Economic Growth. https://doi.org/10.1016/B978-0-08-097086-8.94004-2

Rasmussen, E., & Sørheim, R. (2012). Obtaining early-stage financing for technology entrepreneurship: reassessing the demand-side perspective. Venture Capital: An International Journal of Entrepreneurial Finance. https://doi.org/10.1080/13691066.2012.667908

Harris, A. (2013). Bridging the valley of death. Engineering & Technology. https://doi.org/10.1049/ET.2013.0910

Freear, J., Sohl, J., & Wetzel, W. E. (1994). Angels and non-angels: Are there differences?Journal of Business Venturing. https://doi.org/10.1016/0883-9026(94)90004-3

Macht, S. A., & Robinson, J. (2009). Do business angels benefit their investee companies. International Journal of Entrepreneurial Behaviour & Research. https://doi.org/10.1108/13552550910944575

Sohl, J. E. (2023). THE ANGEL MARKET IN 2023: AN INFLECTION POINT FOR WOMEN ANGELS?Center for Venture Research, 42. https://scholars.unh.edu/cvr/42

Baker, H. K., & Nofsinger, J. R. (2002). Psychological Biases of Investors. Financial Services Review.

Madaan, G., & Singh, S. (2019). An Analysis of Behavioral Biases in Investment Decision-Making. International Journal of Financial Research. https://doi.org/10.5430/IJFR.V10N4P55

Murnieks, C. Y., Haynie, J. M., Wiltbank, R., & Harting, T. (2011). 'I Like How You Think': Similarity as an Interaction Bias in the Investor–Entrepreneur Dyad. Journal of Management Studies. https://doi.org/10.1111/J.1467-6486.2010.00992.X

Armandi, C. (2015). Exploring the drivers of investment decision-making in entrepreneurial pitches.

Klonowski, D. (2018). Screening and Evaluation: Misguided Investigation of Entrepreneurial Firms. https://doi.org/10.1007/978-3-319-70323-7_5

Fisher, G., & Neubert, E. (2022). Evaluating Ventures Fast and Slow: Sensemaking, Intuition, and Deliberation in Entrepreneurial Resource Provision Decisions. Entrepreneurship Theory and Practice. https://doi.org/10.1177/10422587221093291

Steier, L., & Greenwood, R. (1999). Newly created firms and informal angel investors: A four-stage model of network development. Venture Capital: An International Journal of Entrepreneurial Finance. https://doi.org/10.1080/136910699295947

Olaore, R. A., & Adegbesan, S. (2014). The Role of Business Angel in Financing Small Business. Journal of Poverty, Investment and Development.

Maxwell, A. D., Jeffrey, S. A., & Lévesque, M. (2011). Business angel early stage decision making. Journal of Business Venturing. https://doi.org/10.1016/J.JBUSVENT.2009.09.002

Fischer, B. B., Juk, Y., & Avanci, V. de L. (2023, May 14). An Equity, Diversity, and Inclusion Approach in Entrepreneurial and Innovation Ecosystems. https://doi.org/10.55835/6442ffbec93d17c257de1fff

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. The American Economic Review. https://doi.org/10.1257/000282803322655392

Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases.

Wiltbank, R., Sudek, R., & Read, S. (2009). The role of prediction in new venture investing. Frontiers of Entrepreneurship Research.

Cheng, C. X. (2018). Confirmation Bias in Investments. International Journal of Economics and Finance. https://doi.org/10.5539/IJEF.V11N2P50

Payne, J. W., & Bettman, J. R. (2008). Walking with the scarecrow: The information-processing approach to decision research.https://doi.org/10.1002/9780470752937.CH6

Stiensmeier-Pelster, J., & Schürmann, M. (1993). Information Processing in Decision Making under Time Pressure. https://doi.org/10.1007/978-1-4757-6846-6_16

Deci, E. L., & Ryan, R. M. (1985). Information-Processing Theories. https://doi.org/10.1007/978-1-4899-2271-7_8

Simon, H. A. (1956). Rational choice and the structure of the environment. Psychological Review. https://doi.org/10.1037/H0042769

Caplin, A., Dean, M., & Martin, D. (2011). Search and Satisficing. The American Economic Review. https://doi.org/10.1257/AER.101.7.2899

Arvai, J., & Froschauer, A. (2010). Good decisions, bad decisions: the interaction of process and outcome in evaluations of decision quality. Journal of Risk Research. https://doi.org/10.1080/13669871003660767

Hammond, J. S., Keeney, R. L., & Raiffa, H. (1998). Smart Choices: A Practical Guide to Making Better Decisions.

Freear, J., Sohl, J., & Wetzel, W. E. (2002). Angles on angels: Financing technology-based ventures - a historical perspective. Venture Capital: An International Journal of Entrepreneurial Finance. https://doi.org/10.1080/1369106022000024923

Prowse, S. (1998). Angel investors and the market for angel investments. Journal of Banking and Finance. https://doi.org/10.1016/S0378-4266(98)00044-2

Harrison, R. T., Mason, C., & Smith, D. B. (2015). Heuristics, learning and the business angel investment decision-making process. Entrepreneurship and Regional Development. https://doi.org/10.1080/08985626.2015.1066875

Mitteness, C. R., Sudek, R., & Cardon, M. S. (2012). Angel investor characteristics that determine whether perceived passion leads to higher evaluations of funding potential. Journal of Business Venturing. https://doi.org/10.1016/J.JBUSVENT.2011.11.003

Brooks, A. W., Huang, L., Kearney, S. W., & Murray, F. (2014). Investors prefer entrepreneurial ventures pitched by attractive men. Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/PNAS.1321202111

Mason, C., & Stark, J. M. (2004). What do Investors Look for in a Business Plan? A Comparison of the Investment Criteria of Bankers, Venture Capitalists and Business Angels. International Small Business Journal. https://doi.org/10.1177/0266242604042377

Sudek, R. (2006). Angel Investment Criteria. Journal of Small Business Strategy.

Macht, S. A. (2011). Inexpert business angels: how even investors with 'nothing to add' can add value†.Strategic Change. https://doi.org/10.1002/JSC.900

Paul, S., Whittam, G., & Wyper, J. (2007). Towards a model of the business angel investment process. Venture Capital: An International Journal of Entrepreneurial Finance. https://doi.org/10.1080/13691060601185425

Wong, A. Y. (2002). Angel Finance: The Other Venture Capital. Social Science Research Network. https://doi.org/10.2139/SSRN.941228

Gompers, P. A. (1995). Optimal Investment, Monitoring, and the Staging of Venture Capital. Journal of Finance. https://doi.org/10.1111/J.1540-6261.1995.TB05185.X

Sørheim, R. (2003). The pre-investment behaviour of business angels: a social capital approach. Venture Capital: An International Journal of Entrepreneurial Finance. https://doi.org/10.1080/1369106032000152443

Scharfstein, D. S., & Stein, J. C. (1990). Herd Behavior and Investment. The American Economic Review, 80(3), 465–479. https://EconPapers.repec.org/RePEc:aea:aecrev:v:80:y:1990:i:3:p:465-79

Astebro, T. B. (2004). Key success factors for technological entrepreneurs' R&D projects. IEEE Transactions on Engineering Management. https://doi.org/10.1109/TEM.2004.830863

Tversky, A. (1972). Choice by elimination. Journal of Mathematical Psychology. https://doi.org/10.1016/0022-2496(72)90011-9

Maxwell, A. L. (2009). Failing Fast: How and why business angels rapidly reject most investment opportunities. http://hdl.handle.net/10012/4253

Trippi, R. R., & Lee, J. K. (1995). Artificial Intelligence in Finance & Investing: State-of-the-Art Technologies for Securities Selection and Portfolio Management.

Dietterich, T. G. (1996). Machine learning. ACM Computing Surveys. https://doi.org/10.1145/242224.242229

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science. https://doi.org/10.1126/SCIENCE.AAA8415

Lai, T. L., & Xing, H. (2008). Statistical Models and Methods for Financial Markets.

Crowston, K., Allen, E. E., & Heckman, R. (2012). Using natural language processing technology for qualitative data analysis. International Journal of Social Research Methodology. https://doi.org/10.1080/13645579.2011.625764

Schumaker, R. P., & Chen, H. (2006, December 1). Textual Analysis of Stock Market Prediction Using Financial News Articles. Americas Conference on Information Systems.

Liu, Y.-H., Han, T., Ma, S., Zhang, J.-Y., Yang, Y., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhu, D., Li, X., Qiang, N., Liu, T., & Ge, B. (2023). Summary of ChatGPT/GPT-4 Research and Perspective Towards the Future of Large Language Models.arXiv. Org. https://doi.org/10.48550/arXiv.2304.01852

Chiang, C.-H., & Lee, H. (2023). Can Large Language Models Be an Alternative to Human Evaluations?https://doi.org/10.18653/v1/2023.acl-long.870

Röhm, S., Bick, M., & Böckle, M. (2022). The Impact of Artificial Intelligence on the Investment Decision Process in Venture Capital Firms. https://doi.org/10.1007/978-3-031-05643-7_27

Tewari, A., Gabarró, J., Sole, J., Lapouble, B., & Montull, L. (2020). Artificial Intelligence Based Decision Making for Venture Capital Platform. https://doi.org/10.1007/978-3-030-46224-6_11

Tse, R. T. S., Luo, W., D′Addona, S., & Pau, G. (2022). Machine learning-driven credit risk: a systemic review. Neural Computing and Applications. https://doi.org/10.1007/s00521-022-07472-2

Trippi, R. R., & Turban, E. (1992). Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance.

Bahoo, S., Cucculelli, M., Goga, X., & Mondolo, J. (2024). Artificial intelligence in Finance: a comprehensive review through bibliometric and content analysis.4, 1–46. https://doi.org/10.1007/s43546-023-00618-x

Deodhar, O., Bhatkar, S., Dixit, P., Kulkarni, P., Oak, A., & Londhe, S. (2024). Shark Tank Deal Prediction using Machine Learning Techniques.2024 IEEE 9th International Conference for Convergence in Technology (I2CT), 1–6. https://doi.org/10.1109/I2CT61223.2024.10543650

Sherk, T., Tran, M.-T., & Nguyen, T. V. (2019). SharkTank Deal Prediction: Dataset and Computational Model.2019 11th International Conference on Knowledge and Systems Engineering (KSE), 1–6. https://doi.org/10.1109/KSE.2019.8919477

Lavanchy, M., Reichert, P., & Joshi, A. (2022). Blood in the water: An abductive approach to startup valuation on ABC's Shark Tank. Journal of Business Venturing Insights, 17, e00305. https://doi.org/10.1016/j.jbvi.2022.e00305

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Del Ser, J., Bennetot, A., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI.Information Fusion. https://doi.org/10.1016/J.INFFUS.2019.12.012

Jagdale, R., & Deshmukh, M. (2025). Natural Language Processing in Finance: Techniques, Applications, and Future Directions. Machine Learning and Modeling Techniques in Financial Data Science, 411–434. https://doi.org/10.4018/979-8-3693-8186-1.ch016

Osasona, F., Amoo, O. O., Atadoga, A., Abrahams, T. O., Farayola, O. A., & Ayinla, B. S. (2024). Reviewing the ethical implications of ai in decision making processes. https://doi.org/10.51594/ijmer.v6i2.773

Greene, P. G., Brush, C. G., Hart, M. M., & Saparito, P. A. (2001). Patterns of venture capital funding: Is gender a factor?Venture Capital: An International Journal of Entrepreneurial Finance, 3(1), 63–83. https://doi.org/10.1080/13691060118175

Leavy, S. (2018). Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning. 14–16. https://doi.org/10.1145/3195570.3195580

Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., & Criado, N. (2020). Bias and Discrimination in AI: a cross-disciplinary perspective.arXiv: Computers and Society. https://doi.org/10.1109/MTS.2021.3056293

Dey, M. T. (2024). Explainable Artificial Intelligence (XAI). Advances in Environmental Engineering and Green Technologies Book Series, 333–362. https://doi.org/10.4018/979-8-3693-7822-9.ch012

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding Explainability: Towards Social Transparency in AI systems. https://doi.org/https://doi.org/10.1145/3411764.3445188

Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. International Journal of Human-

Computer Studies / International Journal of Man-Machine Studies, 38(3), 475–487. https://doi.org/10.1006/IMMS.1993.1022

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: toward a unified view. Management Information Systems Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. The Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/ANNALS.2018.0057

Choung, H. U., David, P., & Ross, A. (2022). Trust in AI and Its Role in the Acceptance of AI Technologies. International Journal of Human-Computer Interaction, 39, 1727–1739. https://doi.org/10.1080/10447318.2022.2050543

Phillips-Wren, G. (2012). Ai tools in decision making support systems: a review. International Journal on Artificial Intelligence Tools, 21(02), 1240005. https://doi.org/10.1142/S0218213012400052

Buczynski, W., Cuzzolin, F., & Sahakian, B. J. (2021). A review of machine learning experiments in equity investment decision-making: why most published research findings do not live up to their promise in real life. Journal of Data Science, 11(3), 1–22. https://doi.org/10.1007/S41060-021-00245-5

Wetzel, W. E. (1983). Angels and Informal Risk Capital. Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1505193

Goldenberg, J., Han, S., Lehmann, D. R., & Hong, J. W. (2009). The Role of Hubs in the Adoption Processes. Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1960350

Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S. A., & Lomas, E. (2021). Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms. Social Science Research Network. https://doi.org/10.2139/SSRN.3778998

Hernández-Orallo, J. (2017). Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement. Artificial Intelligence Review, 48(3), 397–447. https://doi.org/10.1007/S10462-016-9505-7

Lindsay, G. W. (2020). Attention in Psychology, Neuroscience, and Machine Learning. Frontiers in Computational Neuroscience, 14, 29. https://doi.org/10.3389/FNCOM.2020.00029

Baron, R. A. (2006). Opportunity Recognition as Pattern Recognition: How Entrepreneurs "Connect the Dots" to Identify New Business Opportunities. Academy of Management Perspectives, 20(1), 104–119. https://doi.org/10.5465/AMP.2006.19873412

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1), 4–37. https://doi.org/10.1109/34.824819

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Barnes, N., & Mian, A. (2023). A Comprehensive Overview of Large Language Models. https://export.arxiv.org/pdf/2307.06435v3.pdf

Adomdza, G. (2004). Why do inventors continue when experts say stop? The effects of overconfidence, optimism and illusion of control. http://etd.uwaterloo.ca/etd/gkadomdz2004.pdf

Koziol, K., Schmitz, M., & Bort, S. (2025). Gender differences in entrepreneurial equity financing, a systematic literature review. Small Bus. Econ.https://doi.org/10.1007/s11187-024-00989-x

Arundale, K., & Mason, C. (2024). Business angel groups as collective action: an examination of the due diligence process. Int. Entrep. Manag. J., 21, 39. https://doi.org/10.1007/s11365-024-01052-7

CFA Snapshot. (2015). Canadian Innovation Centre. https://innovationcentre.ca/product-details/cfa-snapshot/

Gonzales M., M. A., Terzidis, O., Lütz, P., & Heblich, B. (2024). Critical decisions at the early stage of start-ups: a systematic literature review. Innov. Entrep., 13, 83. https://doi.org/10.1186/s13731-024-00438-9

Ries, E. (2011). The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. Crown Business.

Mason, C. M., & Harrison, R. T. (2003). Auditioning for money. J. Priv. Equity, 6(2), 29–42. https://doi.org/https://doi.org/10.3905/jpe.2003.320037

Cassar, G. (2009). The financing of business start-ups. J. Bus. Ventur., 19(2), 261–283. https://doi.org/10.1016/s0883-9026(03)00029-6

Baker, H. K., & Ricciardi, V. (2014). How biases affect investor behaviour. Eur. Financial. Rev., 7–10. https://ssrn.com/abstract=2457425

Barbaroux, P. (2014). From market failures to market opportunities: managing innovation under asymmetric information. J. Innov. Entrep., 3(1), 5. https://doi.org/10.1186/2192-5372-3-5

Kliger, D., & Kudryavtsev, A. (2010). The availability heuristic and investors' reaction to company-specific events. J. Behav. Finance., 11(1), 50–65. https://doi.org/https://doi.org/10.1080/15427561003591116

Shark Tank US Dataset. (2024). https://www.kaggle.com/datasets/thirumani/shark-tank-us-dataset

Appendices

Appendix A: Full CFA Rubric from Canadian Innovation Centre

The below questions have been taken directly from the Canadian Innovation Centre's website by permission of Dr. AndrewMaxwell.

1. Features & Benefits (Technology Factor)

Evaluation focus: Demand for your offering and the tangible advantages it delivers against existing solutions.

Assessment question: Does your proposed product or service offer performance advantages compared to currently deployed solutions?

Response levels

A. Substantial advantages over current solutions or no comparable solution yet exists.

B. Measurable benefits over current solutions or clear advantages within a niche.

C. Meets the performance of current solutions at a competitive cost.

Scoring scale

A+ = 10 A = 9 A- = 8 B+ = 6 B = 5 B- = 4 C+ = 2 C = 1 C- = 0 N/A = 0

Guidance & feedback

Early adopters and investors want evidence of measurable benefits. Use low-risk trials, independent testing, or expert endorsements. Map each feature to customer-valued benefits, distinguishing whether your solution is a "pain-killer" (solves an urgent problem) or a "vitamin" (enhances utility). If no clear benefit exists, consider:

- Adapting your business model.
- Bundling into a broader offering.
- Licensing a recognized brand.
- Partnering with strategic suppliers, distributors, or customers.

2. Readiness (Technology Factor)

Evaluation focus: Proximity of the product or service to being market ready.

Assessment question: How far away are you from being able to deliver completed products or services to your first revenue customer?

Response levels

A. Finished offering; beta tests complete; manufacturing/supply issues resolved; possibly initial sales.

B. Tested to expectation; further work needed on approvals or supply chain; no sales yet.

C. Concept complete but further R&D or critical supply-chain work remains.

Scoring scale

A+ = 10 A = 9 A- = 8 B+ = 6 B = 5 B- = 4 C+ = 2 C = 1 C- = 0 N/A = 0

Guidance & feedback

Investors fund businesses, not research projects. Demonstrate that remaining technical hurdles are routine or outsourced to proven partners. Focus on launching with core features; enhancements can follow. Outsource manufacturing where credible partners can scale and solve problems quickly.

3. Barrier to Entry (Technology Factor)

Evaluation focus: Long-term competitive advantage that discourages imitators.

Assessment question: What is unique or patentable about your product that represents a barrier to entry for potential competitors?

Response levels

A. Issued patent, hard-to-replicate proprietary know-how, or unique brand creating significant barrier.

B. Patent pending or specialized know-how; replication is challenging but possible.

C. Innovative product, but limited formal mechanisms to deter future imitators.

Scoring scale

A+ = 10 A = 9 A- = 8 B+ = 6 B = 5 B- = 4 C+ = 2 C = 1 C- = 0 N/A = 0

Guidance & feedback

Patents help but are slow and costly to enforce. Enhance or substitute patents with copyrights, trademarks, licensed IP, economies of scale, exclusive customer or channel agreements, and strategic partnerships that restrict competitor options.

4. Adoption (Market Factor)

Evaluation focus: Evidence that target customers will actually buy.

Assessment question: Can you demonstrate that customers in your target market will purchase your product or service when it is available?

Response levels

A. First customers co-developed the offering and have committed to purchase or trial.

B. Market research confirms demand, but no purchase commitments yet.

C. No independent market validation; no first customers lined up.

Scoring scale

A+ = 10 A = 9 A- = 8 B+ = 6 B = 5 B- = 4 C+ = 2 C = 1 C- = 0 N/A = 0

Guidance & feedback

Involve customers in design, commission third-party validation studies, or obtain endorsements from industry experts who can introduce you to further buyers and investors.

5. Supply Chain (Market Factor)

Evaluation focus: Availability and commitment of suppliers and distribution partners.

Assessment question: Can you confirm there are no success barriers in your supply chain or distribution channel?

Response levels

A. Suppliers/distributors engaged in development and firmly committed to participate at launch.

B. Potential partners identified and interested, but no formal agreements.

C. Partners identified but not yet approached.

Scoring scale

A+ = 10 A = 9 A- = 8 B+ = 6 B = 5 B- = 4 C+ = 2 C = 1 C- = 0 N/A = 0

Guidance & feedback

Unless you produce and sell everything yourself, prove you can secure and manage critical partners. Show partner willingness and your capability to oversee outsourced manufacturing, logistics, or retail promotion.

6. Market Size (Market Factor)

Evaluation focus: Overall market potential and achievable share.

Assessment question: Is the overall market size, and your likely share, sufficient to generate the projected revenues?

Response levels

A. Direct evidence supports a large market; realistic path to revenues exceeding $20 M.

B. Some evidence supports a sizable market; path to revenues exceeding $5 M.

C. No reliable evidence; unclear if revenues can exceed $1 M.

Scoring scale

A+ = 10 A = 9 A- = 8 B+ = 6 B = 5 B- = 4 C+ = 2 C = 1 C- = 0
N/A = 0

Guidance & feedback

Create bottom-up forecasts: customers × price. Use analog businesses, pilot data, or third-party studies for disruptive markets. Demonstrate a focused strategy to capture share, not just a fraction of a vast generic market.

7. Entrepreneur Experience (Management Factor)

Evaluation focus: Team's capability to execute, based on relevant experience.

Assessment question: Do you or members of your team have direct or relevant experience applicable to this business?

Response levels

A. Management has deep, directly relevant experience that spurred creation of the venture.

B. Significant but indirect business or equivalent experience useful to the venture.

C. Primarily technical or limited experience; no direct evidence of operational capability.

Scoring scale

A+ = 10 A = 9 A- = 8 B+ = 6 B = 5 B- = 4 C+ = 2 C = 1 C- = 0
N/A = 0

Guidance & feedback

Highlight prior venture wins, or lessons learned from failures, and complementary team skills. Mitigate gaps with engaged advisers or mentors who have aligned incentives.

8. Financial Expectations (Business-Model Factor)

Evaluation focus: Ability to achieve and manage cash-flow neutrality and eventual profitability.

Assessment question: Do your financial projections persuasively show the company can reach cash-flow neutrality through internal funds, borrowing, and external investment?

Response levels

A. High confidence in reaching cash-flow break-even within 12 months (or clear path to raise funds thereafter).

B. Reasonable confidence in break-even within 24 months.

C. No detailed projection or forecasts show negative cash flow beyond 24 months with uncertain funding.

Scoring scale

A+ = 10 A = 9 A- = 8 B+ = 6 B = 5 B- = 4 C+ = 2 C = 1 C- = 0
N/A = 0

Guidance & feedback

Provide realistic P&L and cash-flow forecasts, stress-test assumptions (price, margin, delay, cost overruns), and outline financing strategy. Identify diverse funding sources and link fundraising rounds to valuation-boosting milestones.

Appendix B: CFA LLM prompts

Below are two samples of the evolution of the Supply Chain prompt that was used for the LLM-based grading of the CFA features. These samples are taken directly out of the prompt JSON that was used for function calling during a pitch evaluation.

V2023-10-3:

"SupplyChain": {

"LetterGrade": "Evaluate the Supply Chain aspect of the startup's product or service and assign an appropriate letter grade (A+, A, A-, B+, B, B-, C+, C, C-, N/A) considering the following criteria:\n\nCan the startup provide confirmation that there are no success barriers regarding their supply chain or distribution channel?\nHave suppliers and channel partners been engaged in the development process, with firm commitments to participate upon market readiness?\nAre there identified possible channel and supply chain partners, even if formal agreements are not yet in place?\nHas the startup approached distribution or supply chain partners at this stage?",

"Evaluation": "Evaluate the Supply Chain aspect of the startup's product or service and provide a comprehensive evaluation, justifying the assigned letter grade. Consider the following points:\n\nThe level of engagement and commitment from suppliers and channel partners in the development process.\nIdentification and initial discussions held with possible channel and supply chain partners.\nThe absence or presence of formal agreements with distribution or supply chain partners.\nAny potential success barriers or challenges related to the supply chain or distribution channel.",

"Recommendations": "Based on the evaluation of the Supply Chain aspect, provide recommendations and suggestions on how the startup can enhance their supply chain management and strengthen their distribution channel. Consider the following aspects:\n\nDemonstrating a good understanding of supply chain requirements and the ability to negotiate with partners.\nDeveloping strong relationships with suppliers, ensuring access to critical raw materials or reliable manufacturing facilities.\nIdentifying and engaging infrastructure partners or distribution channels that can support the startup's product or service.\nCreating a realistic plan to drive traffic to retail locations and foster promotion efforts.\nShowing the startup's importance to distribution partners and their willingness to promote the product/service.\n\nPlease provide specific and actionable recommendations to help the startup optimize their supply chain and distribution channel, ensuring a smooth and efficient flow from production to the market.",

```
  "temperature": 0.2,
  "maxTokens": 7000
 },
 "__v": 0,
 "createdAt": "2023-09-30T01:50:21.759Z",
 "updatedAt": "2023-10-03T19:49:48.302Z"
}
```


V2024-10-06:


```
        "SupplyChain": {
  "LetterGrade": "assign a single grade for the startup's supply chain based on the grading criteria. Firstly, identify if there is a substantial supply chain. If the product or service is completely made by the startup, and they can either sell it direct, or through a well-established channel that has no commercial barrier to entry; in those cases, then they have limited supply chain risks (software as a service, digital content creation, app development, etc). If you identify that the startup is one that has limited supply chain risk, automatically grade the supply chain an A+. However, if this is not the case, then the startup needs to show that they have suitable suppliers and partners available to work with them. The grading criteria is the following. Assign an A+ grade if there is active early involvement of partners with solid commitments and possibly pilot projects or initial orders, showcasing great support for the offering. Assign an A grade if partners are engaged during the development phase, promising to join when the product/service hits the market. Assign an A- grade if partners show interest and have been consulted, yet agreements are somewhat informal or come with certain conditions. Assign a B+ grade if the startup identified potential partners engaged in advanced talks regarding collaborations, indicating substantial interest, albeit without finalized agreements. Assign a B grade if there were initial dialogues held with identified potential partners, who have demonstrated interest, without any formal contracts in place. Assign a B- grade if a roster of possible partners exists, with some initial contact made but substantial discussions pending and interest largely unverified. Assign a C+ grade if identification of possible partners has occurred, along with the creation of strategies for imminent engagement, indicating a clear forward plan despite no actions yet. Assign a C grade if there have been potential partners identified, although yet to be approached. Assign a C- grade if a rudimentary list of potential partners is available, lacking specific plans or strategies for engagement. Assign an NA grade if there is no data in the pitch to analyze the supply chain or the extent of partner interaction.",
  "Evaluation": "Provide a comprehensive and cohesive paragraph evaluating the startup’s supply chain. This factor is about demonstrating that there are no success barriers either about the supply chain or distribution channel. Begin by identifying its basic supply chain needs: a comprehensive network involving various channel partners, or a direct sales route without commercial barriers. Next, delineate the current phase of its supply chain structure - embryonic with initial plans, evolving with ongoing discussions, or advanced with firm commitments and active engagements. Delve into a detailed analysis of the evidence provided or implied in the pitch, highlighting the robustness of the supply chain infrastructure, including the extent of partnership with
```

channel partners and strategies to engage suppliers, along with approaches to mitigate potential risks. Evaluate these aspects to ascertain the startup's potential in establishing a resilient and efficient supply chain capable of fostering sustained growth. If the startup utilizes a direct sales or well-established channel, analyze both the supply chain and distribution facets focusing on significant aspects like digital network efficiency, production workflow, customer engagement, direct marketing strategies, and quality control, among others. Critically assess the most relevant aspects pertaining to the startup's current operations and future growth. Maintain a professional and solutions-focused tone, emphasizing constructive analysis while adhering to an inquisitive approach to explore the full potential of the startup's supply chain aspects.",

"Recommendations": "provide a comprehensive and cohesive paragraph with specific, actionable, tailored, and revolutionary recommendations intended to bolster the startup's supply chain. ",

"temperature": 0.4,

"maxTokens": 7000

},

"__v": 0,

"createdAt": "2023-10-03T20:22:11.283Z",

"updatedAt": "2024-10-06T21:52:29.190Z"

}

Example pitches

Example 1:

"Hi, Sharks. My name is Xin Wang. And my name's Melissa Colella-Wang, and we're from New York, seeking $300,000 for 3% equity of our company.

Sharks, let's talk about something important... Tacos. Everyone loves tacos, but my doctors say I can't eat tacos every day. That's 'cause of the calorie-laden tortilla. And switching to lettuce wraps? Uh, no. They break apart whenever you pick them up, and they just don't taste great. That combined If only there was something of a lettuce wrap the nutritional benefit of a traditional tortilla. And the functionality Well, now there is. Say hello to... Jica wraps! Jica wraps! A crisp and refreshing tortilla alternative of fresh jicama. Made from thin slices Jica what?! Corcoran: Oh! Jica yeah! Hey! It's Jicaman, a member of the League of Superfoods! Jicama's our hero ingredient 'cause it truly is a superfood. It's super low calorie... ...super low carb... super high fiber... and one of its

superpowers is improving gut health.Ooh! Jica yeah!

Xin: Compared to a tortilla, which contains 100 calories and 15 carbs, our wraps are only eight calories and one net carb per wrap. Use them to "skinny" any taco. Round and pliable, Plus, they're perfectly and their neutral taste pairs well with any recipe. Paleo, keto, So whether you're gluten-free, Whole30, or just looking for a healthier way to make every day Taco Tuesday... you can say "See ya, tortilla." "Lettuce do better with... [Let us] Jica wraps!" Jica yeah! Come on, Sharks, and join the jica mania.

Jica what?! Xin: Jica what?! So you guys have got samples of our Jicama Tacos, you've got samples of our Jicama Sticks with cilantro-lime seasoning, and our Jica de Gallo. Pico de gallo with jicama in The Jica de Gallo is our it. Without the stuffing? Can I get one I just want to taste the actual tortilla. Jicama before? Paltrow: Have you never had very pleasant-tasting... It's like a watery, I have not had it before. Slightly sweet root. I actually have jicama taco recipes in my last cookbook. Greiner: Oh, you do? How much do you sell it for? What does it cost to make?

So, we make it for $1.85. $2.99, and it's on shelf for We sell it to retail for $4.99. And how many do I get in that package? So you have about 15 wraps per package. About $0.30. Each wrap is actually a regular tortilla like this. You're gonna pay $0.28 for like a cassava And a grain-free tortilla or an almond-flour tortilla, you're paying over a dollar a lot of times for that kind of a tortilla. So it's very reasonably priced.

Paltrow: I see, as a trend in the wellness world, a huge thing. Grain-free is becoming First, it was gluten-free and now it's, like, full on grain-free. So people are not even having rice or quinoa. This category. They're really moving into getting traction, for sure. So I understand why it's Hey, are you making any money? Do you have any profits?

So, we started the company in 2015, and we launched with a different product. It was a jicama chip. And that jicama chip, we got to $500,000 in sales. The issue was, we were

competing in the salty-snacks category, and we were competing with large giants, and we were losing so much money. We got to the point where we only had $300 in our bank account and $70,000 in credit-card debt. We literally Airbnb'ed our house just to survive. Oh, my gosh. We sold our car. We sold everything. So we pivoted in 2019 from our jicama chips to fresh-cut jicama. So in 2019, just one year sales, $1 million. Wow! 2020, $4 million. Corcoran: Wow. 2021, $5 million. What?! This year, we're doing $6 million. Wow. Okay. This year, in terms of profits, about $400,000, about the same as last year. Down this year Our profits are a little because we've been putting a lot of our money and resources into actually building our own self-manufacturing capacity. Where are you manufacturing?

The jicama products are processed and sourced in Mexico. Mexico's actually the only country that grows jicama commercially. So we like to put our manufacturing very close to the source. Paltrow: And how long does it stay fresh? I mean, they package it in Mexico and ship it? Is 19 days, Yeah, so, our shelf life which is actually 2 to 3 times processors what normal, fresh-cut will get on Jicama optimized the packaging because we've really to extend that shelf life. What national retailers are you in, or grocers?

Some of the retailers that we're in are Whole Foods, Sprouts, and we actually have a big push into food services here with Sysco. We're actually being pushed into all of their distribution centers. So that's a great thing for food services. So how many stores are you in?

We're in about 700 doors. And that leads to $6 million in sales? They're very popular."

Example 2:

"Hi Sharks My name is Amrita Saigal, Hi, Sharks. I am from Los Angeles, California. and I am seeking $250,000 for 5% equity in my company, Kudos.

So, Sharks, this may seem a little odd to you, being wrapped in plastic, disposable

diapers are made but this is what most of, meaning most babies spend two to three years of their lives wrapped in plastic, not letting their butts breathe, and not to mention, diapers are the third-largest consumer item in landfills. So we decided to diaper the world differently with Kudos.

Kudos are premium disposable diapers where baby's sensitive tush touches 100% soft, natural cotton all day... Not plastic. Cotton is the number-one doctor-recommended material for rash and eczema. It's natural, chlorine free, and far more sustainable. And these are natural diapers that actually work! Our team of MIT engineers developed award-winning, double-dry technology, patent-pending, absorption layers that uses two instead of just one to avoid those leaks and blowouts and keep your baby sleeping all night long.

So, Sharks, who's in to help change the world with us one tush at a time? So, in front of you, you each have an individual pack of diapers. The first thing you might notice about them is how soft they are. Our customers love about them. And this is something Greiner: So very soft, like, if I'm comparing, right? Saigal: Yeah. The plasticky feel. Doesn't have Obviously there's something on the outside here that keeps it leakproof. Yeah. Not super thick. Correct. Your average diaper, for sure. I think this is thinner than Yep. Will you tell us a little about your background?

So, I come from a family of all engineers. I was fortunate to go to MIT, in mechanical engineering, where I majored and I later went to Harvard to get my MBA. I got my dream job out of college, engineer for Procter & Gamble. Working as a manufacturing And I saw firsthand just how much plastic sanitary pads and diapers. we were putting into to take a huge risk. So I decided I decided to quit my stable job and move to India and start my first company there, making biodegradable sanitary pads banana-tree fiber, out of waste and that has grown to become eco-friendly brands there. one of the leading I transitioned out of that, moved back to the US, and as I entered my early 30s, my first wave of friends started having babies, and I knew we could do better with diapers, as

well. Do you have a baby at this time, or just for your friends? No. No. Um, yeah, no baby. I actually am now a new mom with a 12-week-old at home. Congrats. Corcoran: Ah. There's a picture of her right here where she is wearing Kudos and has been testing them every day. Kind of wild. I actually got the call for... from the producers for "Shark Tank" when I was in the delivery room, and I was like, "Oh, my God! That's crazy. How is this all gonna happen in the next couple of weeks?" That's a very special day, then, huh? A lot of new beginnings. A lot of new beginnings. Saigal:

But, like, as a new mom, this is even more personal. I want the cleanest and safest ingredients touching her... her skin. And what are your sales? So, we launched a year ago. We have done $850,000 in In this first year, sales. That's pretty good. Greiner: It's very good. O'Leary: And how have you sold... direct to consumer or you're going through retail distribution? We are... been 100% direct to consumer. Have you bootstrapped this whole thing? We have not. We have raised $3.2 million. Dang! Wow. O'Leary: And what valuation? Our last valuation was $12.5 million. I'm assuming you're not profitable. Saigal: We are not profitable. What are your aggregate losses since you started this? Since we've started, $1.5 million. That being said, we do expect to become profitable next year. $4.5 million sales next year. We do expect to do How much are the diapers? Saigal: We have two options. A one-time purchase for $88 You can make or join our subscription for $78 a month. Each box of diapers we sell is a one-month supply of diapers. How does your product price compared to the general market out there? So, in the eco-friendly space, we are very much on par with them. Are the eco-friendly category of diapers about twice as expensive as the rest of the diapers? Not twice as expensive. I would say it's anywhere from 20% to 50%. And what is your margin on the diapers? Saigal: Our margin is $39. So, landed at our warehouse That's not a lot of margin for a direct-to-consumer business. For the diaper industry, And the reality is, it's very, very typical. No, I'm not saying this, but you're direct to consumer. How much do you spend on marketing? $100,000 in marketing in our We have spent less than first year. Of customer acquisition Our blended cost has been $23, which we're really proud of. has been huge for us. Word of mouth So where do you see yourself in five years? I really think of baby

diapers as just the beginning. We just launched with baby wipes three weeks ago. And then, additionally, with training pants and swim pants. That's just the beginning. Then where I really see us going, Mark, is the adult diaper market. You could be a customer. So, Kevin, It is a booming market. Saigal: Listen, I'd like to wear a pair with the Mavs logo on it. This is a very mature market. You worked at the largest player in this space. I did. There was a behemoth brand there. Yeah. Do you really think, after you get, let's say, 4% or 5% share of the market, they're not gonna wake up and say "We need an eco-friendly diaper"? And I know you've said it's very hard to do... Kevin, if they get 4%... If she gets 4% or 5%, she doesn't care. The numbers into perspective, Kevin, just to put 0.01% of the market would be $7 million in sales. Paltrow: But these big incumbents, they cannot build brands themselves in-house. You see it all the time. Companies are... I mean, big soft-drink Gwyneth, the only way you get out... An authentic brand. They can't develop So anyway, let me... While they're arguing, right? What are you missing? In terms of what? What are you looking for? We need to figure out how to make this product a household name. We know we have a better product, we know parents love it. We need to expand into other online retailers."

Human vs LLM CFA Grading Results

Below are the 31 pitches used to compare the correlation between the prompted-LLMs CFA scores and the Human Consensus CFA scores.

Table 10. Human vs. LLM CFA Grading Results for 31 Pitches

Pitch #	Outcome (0 = no deal, 1 = deal)	Human Consensus Total	GPT-4 Total	4.1-mini Total	O3 Total	4.1 Total
1	1	55.67	64	61	59	70
2	0	41.67	56	48	50	47
3	1	56.67	57	45	49	53
4	0	47.33	51	53	44	46
5	1	54.33	61	52	49	61
6	1	64.33	72	71	68	73
7	0	45.33	64	46	48	62
8	1	51.33	56	55	48	69
9	0	49.67	50	50	46	48
10	0	51	61	55	46	61
11	1	59.67	62	64	53	57
12	1	40.67	27	37	28	37
13	0	40.33	34	42	34	38
14	1	50	51	54	51	64
15	1	58.67	71	65	64	70
16	1	56.67	44	57	48	47
17	1	42	44	45	39	50
18	0	41	48	40	39	45
19	0	39.67	34	27	38	35
20	1	48.33	55	58	41	48
21	1	59	68	67	56	64
22	1	53	72	54	54	66
23	0	46.33	41	47	33	38
24	1	56.67	71	65	58	67
25	0	60.67	59	63	56	67
26	1	47.67	53	54	51	60
27	1	39.67	36	44	38	49
28	0	43	58	42	46	55
29	1	45.33	37	44	38	39
30	0	60.33	73	67	55	73

31	0	37.33	31	39	35	34

Example Evaluation and Recommendation from Unprompted GPT4 vs Prompted and Engineered GPT4

Unprompted GPT4 Output

Reasons for Recommendation:

1. Strong Market Need & Differentiation: The problem Amrita Saigal addresses is significant, both environmentally and in terms of consumer health, diapers are a major landfill component, and traditional offerings are mostly plastic-based. Kudos differentiates itself by offering a cotton-based solution that is both eco-friendly and skin-friendly, targeting concerns like rashes and eczema, which are common in infants.
2. ***Innovative Technology:*** *The double-dry technology developed by a team of MIT engineers provides a strong unique selling proposition. The patent-pending status may offer a competitive edge in a market dominated by traditional players.*
3. ***Market Validation & Growth Potential:*** *The sales figures ($850,000 in the first year) indicate a robust market acceptance and potential for growth. The strategic expansion into adjacent products like baby wipes and plans for entering the adult diaper market demonstrate foresight and scalability.*
4. ***Founder's Background and Commitment:*** *Amrita's background in engineering and experience at Procter & Gamble, coupled with her entrepreneurial success in India, positions her strongly to lead this venture. Her personal story as a new mother adds authenticity and a deeper understanding of the consumer experience.*
5. ***Financials and Valuation:*** *While the company is currently not profitable with aggregate losses of $1.5 million, the projections to turn profitable next year with expected sales of $4.5 million are promising. The company's valuation at $12.5 million with prior funding of $3.2 million suggests confidence from previous*

investors, though the current asking valuation might need scrutiny given the financials.

Risks:

- ***Market Penetration and Competition:*** *Breaking into a market with established players will require significant marketing and brand positioning efforts. The risk of big players moving into the eco-friendly segment cannot be underestimated.*
- ***Cost and Margins:*** *The margins are relatively tight for a direct-to-consumer model, which could challenge scalability and profitability, especially with planned expansion and product development needs.*

Advice for the Entrepreneur:

1. ***Enhance Branding and Consumer Education:*** *Focus on educating consumers about the benefits of natural cotton diapers versus traditional plastic ones. Leveraging digital marketing and storytelling can be powerful, especially stories that resonate with new parents' concerns about health and sustainability.*
2. ***Explore Strategic Partnerships:*** *Collaborating with pediatricians, dermatologists, and eco-friendly marketplaces can increase credibility and market reach. Partnerships with baby product influencers and eco-conscious community leaders could also drive brand awareness and acceptance.*
3. ***Cost Management:*** *Investigate ways to optimize production and supply chain to improve margins. As the company scales, there might be opportunities to negotiate better terms with suppliers or streamline logistics.*
4. ***Expand Distribution Channels:*** *While direct-to-consumer has proven effective, expanding into online and offline retail could significantly increase visibility and sales. This could include eco-friendly stores, large retailers with a natural product line, and specialty baby stores.*
5. ***Prepare for Competitive Response:*** *Develop a contingency plan for potential market entry by larger incumbents. This could involve accelerating brand loyalty*

programs, enhancing product offerings, or securing intellectual property rights that can offer a longer-term competitive advantage.

Overall, Kudos presents an attractive investment opportunity with its innovative product and strong market positioning. However, careful attention to financial management, competitive strategy, and scaling operations will be crucial for long-term success.

Prompted and Engineered GPT4 Response:

Factor: Features and Benefits
Grade: A-
Factor overview: Does your proposed product or service offer performance advantages compared to currently deployed solutions?
Evaluation: Kudos distinguishes itself in the disposable diaper market through its use of 100% soft, natural cotton that replaces the traditional plastic, addressing both environmental concerns and skin sensitivity issues such as rashes and eczema. The product's patented double-dry technology, developed by MIT engineers, offers superior absorption with two layers to prevent leaks and blowouts, enhancing usability and reliability. This innovation, combined with natural, chlorine-free materials, positions Kudos as a premium, eco-friendly alternative to conventional diapers, aligning well with increasing consumer demand for sustainable baby products. The pitch evidences substantial market traction with $850,000 in sales within the first year, a strong direct-to-consumer model, and a growing customer base that values the softness and effectiveness of the product. The founder's background in engineering and prior success with eco-friendly sanitary products further supports the company's capability to deliver on its promises. Additionally, the expansion into wipes and plans for training and swim pants, as well as adult diapers, indicate a strategic vision for market growth. While the product competes in a niche eco-friendly segment, it offers clear advantages over traditional diapers in terms of sustainability, skin health, and leak prevention, which are significant benefits that resonate with modern parents. The company's ability to maintain competitive pricing within the eco-friendly category and manage customer acquisition costs effectively underscores its operational viability and potential for scaling.
Recommendation: To enhance Kudos' value proposition, the company should focus on broadening its retail presence beyond direct-to-consumer channels to increase brand visibility and accessibility, which is critical for becoming a household name. Investing in targeted marketing campaigns that highlight the unique double-dry technology and the

health benefits of natural cotton could deepen consumer engagement and differentiate the brand further from competitors. Exploring partnerships with pediatricians and eco-conscious parenting communities could build trust and advocacy. Additionally, accelerating product line expansion into complementary baby care items and adult diapers with eco-friendly features could capture a larger market share and diversify revenue streams. Finally, ongoing innovation to improve cost efficiency and margins without compromising product quality will be essential to sustain growth and profitability in a competitive market.

Factor: Readiness

Grade: A

Factor overview: How far away are you from being able to deliver completed products or services to your first revenue customer?

Evaluation: Kudos is in the late stage of readiness, characterized by a fully developed and market-tested product with significant sales traction. The startup has demonstrated strong market validation with $850,000 in sales within the first year, a growing customer base, and positive customer feedback highlighting the product's unique value proposition of natural cotton diapers with patented double-dry technology. The product is complete and has been tested extensively by real customers, including the founder's own child, which adds credibility. Operational readiness is evident from the established direct-to-consumer sales channel and a well-managed supply chain, despite the typical low margins in the diaper industry. Financially, Kudos has raised $3.2 million with a $12.5 million valuation, indicating investor confidence and financial stability to support growth and expansion. The company is actively expanding its product line with baby wipes and plans to enter adjacent markets like training pants and adult diapers, signaling a strategic growth trajectory. Overall, Kudos shows strong readiness for scaling and market expansion, with minor challenges around profitability and marketing spend that are common at this stage.

Recommendation: To accelerate market entry and growth, Kudos should focus on scaling its marketing efforts to build brand awareness beyond its current direct-to-consumer

base, potentially leveraging partnerships with online retailers and eco-friendly marketplaces to broaden distribution. Enhancing supply chain efficiencies to improve margins will be critical, as will continuing to gather and showcase customer testimonials and clinical endorsements to strengthen trust and differentiation. Exploring strategic collaborations with pediatricians and eco-conscious parenting communities could amplify credibility and adoption. Additionally, investing in data analytics to optimize customer acquisition costs and subscription retention will support sustainable growth. Finally, preparing for entry into adjacent markets like adult diapers with targeted R&D and market research will position Kudos for long-term expansion and diversification.

Factor: Barrier To Entry

Grade: A-

Factor overview: What is unique or patentable about your product that represents a barrier to entry for potential competitors?

Evaluation: Kudos presents a meaningful barrier to entry primarily through its patent-pending double-dry technology, which employs two absorption layers to prevent leaks and blowouts, a feature that is not commonly found in standard disposable diapers. This proprietary technology, developed by a team of MIT engineers, offers a distinctive product advantage that is difficult for competitors to replicate quickly without infringing on intellectual property rights. Additionally, the focus on 100% natural cotton as the diaper's contact material addresses a significant market demand for safer, eco-friendly baby products, which enhances brand differentiation and customer loyalty. However, while the patent-pending status provides some protection, the market is mature and dominated by large incumbents with deep resources who could potentially develop similar eco-friendly products or acquire technology to compete. The brand's early traction, direct-to-consumer model, and planned expansion into adjacent markets like baby wipes and adult diapers provide some strategic moat, but these barriers are moderate given the potential for new entrants or incumbents to innovate or leverage scale. Overall, Kudos has established a solid foundation of barriers through technology and brand positioning, but these must be fortified to maintain a sustainable competitive

advantage in the long term.

Recommendation: To strengthen Kudos' barriers to entry immediately and sustainably, the company should prioritize securing full patents for its double-dry absorption technology to legally protect its unique product features and deter direct replication. Investing in continuous R&D to improve the technology and develop additional proprietary innovations, such as advanced bio-based materials or enhanced absorption mechanisms, will create a technology moat that evolves beyond current capabilities. Expanding the brand's presence through strategic partnerships with major online retailers and leveraging customer testimonials to build an authentic, trusted brand can increase customer loyalty and reduce price sensitivity. Additionally, exploring proprietary supply chain advantages, such as exclusive sourcing agreements for natural cotton or sustainable materials, can raise operational barriers for competitors. In the longer term, developing formal mechanics like subscription models with high customer retention, data-driven personalization, and community engagement programs will embed customer stickiness. Finally, exploring adjacent markets such as adult incontinence products with eco-friendly features can diversify revenue streams and capitalize on the brand's trust and technology, creating layered barriers that make it difficult for competitors to match Kudos' comprehensive offering.

Factor: Adoption

Grade: A

Factor overview: Can you demonstrate that customers in your target market will purchase your product or service when it is available?

Evaluation: Kudos is currently in a mature phase of adoption, demonstrating substantial evidence of customer engagement and market traction. The startup has achieved $850,000 in sales within its first year, operating primarily through a direct-to-consumer model, which indicates a validated demand for its eco-friendly disposable diapers. The founder's background and personal connection to the product add credibility and authenticity, which likely resonates well with the target market of environmentally conscious parents. Customers have responded positively to the product's softness and

natural materials, as highlighted by direct feedback during the pitch. The company has also raised $3.2 million in funding, suggesting investor confidence in the product's market potential. Marketing efforts have been efficient, with a customer acquisition cost of $23 and significant word-of-mouth growth, further signaling strong adoption. The introduction of complementary products like baby wipes and plans to expand into training pants and adult diapers indicate a strategic approach to broadening the customer base and increasing retention. While profitability has not yet been achieved, projections for $4.5 million in sales next year demonstrate a positive growth trajectory. The pitch also reveals awareness of competitive pressures from established brands but emphasizes the difficulty incumbents face in authentically entering the eco-friendly niche, positioning Kudos well for continued adoption and market share expansion. Overall, the evidence points to a well-validated product with enthusiastic early adopters and a clear path for scaling adoption through product line expansion and increased brand recognition.

Recommendation: To elevate Kudos' market adoption further, the startup should focus on scaling its marketing efforts to transform brand awareness into a household name, leveraging the strong word-of-mouth momentum and customer testimonials as social proof. Expanding partnerships with online retailers and exploring selective retail distribution could increase accessibility and convenience for customers, driving higher sales volumes. Implementing a structured customer feedback loop will enable continuous product refinement and foster deeper customer involvement, potentially turning users into brand advocates who contribute to product development. Additionally, investing in targeted influencer marketing within the parenting and eco-conscious communities could amplify reach and credibility. Exploring subscription model enhancements, such as personalized offerings or loyalty programs, may improve customer retention and lifetime value. Finally, preparing for competitive responses by emphasizing Kudos' unique patented technology and sustainability credentials will help maintain differentiation as the market evolves.

Factor: Supply Chain

Grade: B+

Factor overview: Can you provide confirmation that there are no success barriers either about your supply chain, or distribution channel?

Evaluation: Kudos operates with a direct-to-consumer sales model, which inherently reduces supply chain complexity and commercial barriers typically associated with multi-channel distribution. The startup's supply chain needs focus primarily on sourcing high-quality, natural cotton and manufacturing diapers with their proprietary double-dry technology. The pitch indicates that Kudos has successfully launched and generated significant revenue ($850,000 in the first year) with a direct sales strategy, supported by a digital platform for customer engagement and subscription services, which suggests an efficient production workflow and customer acquisition approach. However, the pitch lacks explicit details about supplier partnerships, manufacturing scale, or risk mitigation strategies, which are critical for sustaining growth given the product's physical nature and reliance on raw materials. The supply chain appears to be in an evolving phase, with evidence of operational execution but limited transparency on the robustness of supplier relationships or contingency plans. The company's marketing spend and customer acquisition cost indicate a lean approach, but scaling will require fortified supply chain partnerships and possibly expanded distribution channels, including online retailers as mentioned. Overall, Kudos shows promising direct-to-consumer supply chain management but would benefit from deeper integration with suppliers and strategic partners to ensure resilience and scalability.

Recommendation: To strengthen Kudos' supply chain, the startup should formalize partnerships with key suppliers and manufacturers to secure raw material availability and production capacity, potentially through pilot projects or initial purchase commitments that demonstrate mutual confidence. Establishing these firm commitments will mitigate risks related to supply disruptions and cost volatility. Additionally, expanding distribution beyond direct-to-consumer to include selective online retailers will require developing robust logistics and inventory management systems to maintain product quality and delivery reliability. Investing in supply chain transparency and

traceability, especially given the eco-friendly positioning, can enhance brand trust and customer loyalty. Finally, implementing advanced forecasting tools and flexible manufacturing processes will allow Kudos to adapt swiftly to market demand fluctuations, supporting sustainable growth and profitability.

Factor: Market Size

Grade: A-

Factor overview: Is the overall size of the market and your likely market share, sufficient to generate the envisaged revenues? Further, is the overall market forecast to be large enough to be interesting?

Evaluation: Kudos has demonstrated a strong initial market presence with $850,000 in sales within its first year, operating entirely direct-to-consumer and having raised $3.2 million in funding with a valuation of $12.5 million. Their sales target of $4.5 million for the next year indicates a clear trajectory toward scaling. The pitch references the broader diaper market, highlighting that even a 0.01% market share equates to $7 million in sales, suggesting a market potential well above $20 million. This is supported by the founder's industry experience at Procter & Gamble and the innovative product features like the patent-pending double-dry technology and eco-friendly materials, which align with growing consumer demand for sustainable baby products. However, the evidence backing these projections is primarily internal sales data and qualitative market insights rather than authoritative third-party market research or extensive primary data. The strategy to expand product lines into baby wipes, training pants, swim pants, and eventually adult diapers indicates a plan to leverage brand strength and diversify revenue streams, which could realistically support reaching and surpassing $20 million in sales. Risks include competition from large incumbents and the challenge of scaling brand awareness beyond direct-to-consumer channels, as acknowledged by the founder's desire to expand into other online retailers. Overall, the startup presents a promising market opportunity with a feasible growth plan but would benefit from more explicit, data-backed market validation to fully substantiate its ambitious sales targets.

Recommendation: To elevate Kudos' market potential and confidently surpass the $20

million sales target, the startup should invest in comprehensive third-party market research and in-depth primary consumer studies to substantiate market size and customer demand more explicitly. Enhancing marketing efforts with innovative digital campaigns and strategic partnerships with major online retailers will amplify brand visibility and accelerate customer acquisition beyond the current direct-to-consumer model. Expanding product offerings thoughtfully while maintaining product quality and sustainability will help capture a broader share of the baby care market and facilitate entry into the lucrative adult diaper segment. The company should also develop agile strategies to anticipate and respond swiftly to competitive moves from incumbents, leveraging its authentic brand story and unique technology. By combining rigorous data-driven market validation with dynamic growth tactics and diversified product lines, Kudos can solidify its position as a household name and unlock substantial long-term market expansion.

Factor: Entrepreneur Experience

Grade: A

Factor overview: Do you or members of your team have any direct or relevant (entrepreneurial, industry, business) experience that can be directly applied to the challenges facing this business?

Evaluation: The entrepreneurial experience of the startup's founder, Amrita Saigal, is notably strong and highly relevant to the business challenges faced by Kudos. Saigal's background includes a mechanical engineering degree from MIT and an MBA from Harvard, equipping her with both technical and business acumen. Her direct industry experience is significant, having worked as a manufacturing engineer at Procter & Gamble, a major player in the diaper and sanitary products market. This role provided her with firsthand insight into the environmental impact and manufacturing challenges of disposable hygiene products. Furthermore, Saigal has prior entrepreneurial experience, having founded and grown a successful eco-friendly sanitary pad company in India, demonstrating her ability to lead and scale a business in a related eco-conscious product space. This blend of technical expertise, corporate experience in the relevant industry,

and proven entrepreneurial success forms a solid foundation for Kudos. Additionally, her recent transition into motherhood adds a personal and authentic connection to the product, likely enhancing her understanding of customer needs. While the pitch does not detail the broader management team's composition, Saigal's comprehensive experience and strategic vision suggest a strong leadership core. Potential gaps might exist in scaling operations or retail distribution expertise, which could be addressed through strategic hires or advisors. Overall, the founder's deep and relevant experience, combined with her demonstrated ability to innovate and lead in the eco-friendly hygiene market, positions Kudos well for success.

Recommendation: To further strengthen the startup's entrepreneurial foundation, it is recommended that the team actively recruit or consult with seasoned experts in retail distribution and supply chain scaling to complement Saigal's strong product and manufacturing background. Establishing a robust advisory board with veterans from the consumer packaged goods (CPG) industry could provide critical insights and open doors to strategic partnerships. Additionally, leveraging Saigal's unique blend of technical and business expertise, the company should invest in leadership development programs to build a well-rounded executive team capable of managing rapid growth. To capitalize on the founder's authentic connection to the target market, the startup could deepen customer engagement through community-building initiatives and user-generated content, enhancing brand loyalty. Finally, documenting and communicating the founder's entrepreneurial journey and industry insights more prominently in marketing and investor relations could further build credibility and attract strategic investors or partners.

Factor: Financial Expectation

Grade: A-

Factor overview: Do your financial projections present a persuasive argument that your company can achieve cash-flow neutrality, based on your own investment, money you can borrow, and money you can raise from external investors?

Evaluation: The startup, Kudos, presents a compelling financial narrative with $850,000

in sales in its first year and a substantial $3.2 million raised to date, reflecting strong investor confidence and market validation. Their projection to reach $4.5 million in sales next year and achieve profitability demonstrates a clear and ambitious path to cash-flow neutrality. The $12.5 million valuation and $1.5 million aggregate losses indicate prudent scaling with a focus on growth over immediate profitability. The direct-to-consumer model with a $39 margin per diaper box, coupled with a relatively low customer acquisition cost of $23, suggests a viable unit economics framework. However, the financial projections rely on optimistic assumptions about market penetration and sales growth, and there is limited detail on cash flow management strategies or contingency plans for potential shortfalls. While the company has secured significant funding and demonstrated early sales traction, further clarity on milestone-driven financing rounds and detailed cash flow forecasts would strengthen confidence in their financial expectations. Overall, Kudos shows a realistic and promising route to cash flow neutrality within 12 months, supported by credible evidence and a strong growth trajectory, albeit with some risks inherent in scaling and market competition. Recommendation: To fortify Kudos' financial backbone, the company should develop and publicly share detailed, milestone-based financial roadmaps that align fundraising rounds with specific achievements such as sales targets, patent approvals, or market expansions. Enhancing transparency around cash flow management and establishing conservative contingency plans will mitigate risks associated with optimistic sales projections. Additionally, diversifying funding sources and exploring strategic partnerships or retail expansions could provide alternative cash inflows and reduce dependency on a single revenue stream. Implementing robust financial modeling tools to regularly update forecasts based on real-time data will enable agile decision-making and investor confidence. Finally, emphasizing sustainable growth over rapid scaling will help preserve cash reserves and position Kudos for long-term profitability and market leadership in the eco-friendly diaper segment.

Example Code Snippets from Classification Modeling

Model Evaluation Metrics

"Specificity_scorer" creates a metric that evaluates the model's specificity from its confusion matrix. "balanced_score" is a metric we introduced to identify top performing models during evaluation. We introduced specificity to ensure our model captures True Negatives (pitches that did not receive investment).

Figure 3. Model Evaluation Metrics

```
def specificity_scorer(y_true, y_pred):
    tn, fp, fn, tp = confusion_matrix(y_true, y_pred).ravel()
    return tn / (tn + fp) if (tn + fp) else 0

def balanced_score(y_true, y_pred):
    f1   = f1_score(y_true, y_pred, average='binary')
    acc  = accuracy_score(y_true, y_pred)
    spec = specificity_scorer(y_true, y_pred)
    return (f1 + acc + spec) / 3
```

Hypothesis 2a Testing

Sample code that took the AI model's results of hypothesis 2 and compared them to the results of the 3 expert graders.

Figure 4. Hypothesis 2a Testing

```
def test_on_hypothesis2(model, features, scaler, sampler):
    test_data = pd.read_excel('Hypothesis2.xlsx')
    ...
    pitch_comparison = pd.DataFrame({
        'Pitch_ID': ...,
        'Actual_Deal': y_test.values,
        'Predicted_Deal': y_pred,
        'Prediction_Probability': y_prob,
        'Prediction_Type': [...]
    })
    pitch_comparison.to_csv('grid_results/hypothesis2_pitch_by_pitch_results.csv')
```

Feature Importance and Model Explainability

Figure 5. Feature Importance and Model Explainability

```
rf.fit(X_train[features], y_train)
gb.fit(X_train[features], y_train)
lr.fit(X_train[features], y_train)
importance_df['Average_Importance'] = \
    importance_df[['RF_Importance','GB_Importance','LR_Importance']].mean(axis=1)
...
if shap_available:
    explainer = shap.TreeExplainer(model) \
        if isinstance(model,(RandomForestClassifier,GradientBoostingClassifier)) \
        else shap.KernelExplainer(model.predict_proba, X_train)
    shap_values = explainer.shap_values(X_test)
```

Final and Isolated Setup of Best Performing Model (GPT4)

Below is the final model configuration that yielded the best predicitive results.

Code Output for Best Performing Model

Below is the output from the best performing model:

Soft Voting Ensemble – Naive Bayes, Logistic Regression, Random Forest, Gradient Boosting (GPT-4). Using features: Factor 1,3,8, Total CFA score, Ask $, Ask %

Detailed Model Report

1. Model Configuration

Table 11. Best-Performing Model Configuration

Parameter	Value
Model Type	voting_soft_2
Feature Set	factors_1_3_8
Scaler	standard
Sampler	tomek
# of Features	6

2. Performance Metrics

Table 12. Best-Performing Model Performance Metrics

Metric	Score
Accuracy	0.850
F1 Score (Macro)	0.833
Precision (Macro)	0.830
Recall (Macro)	0.838
Specificity	0.800
ROC AUC (Macro)	0.816
Avg. Precision	0.887
Balanced Score	0.828

3. Confusion Matrix

Table 13. Best-Performing Model Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative	16	4
Actual Positive	5	35

Counts:

- True Negatives (TN): 16
- False Positives (FP): 4

- False Negatives (FN): 5
- True Positives (TP): 35

4. Classification Report

Table 14. Best-Performing Model Classification Report

Class Label	Precision	Recall	F1-score
0 (Negative)	0.762	0.800	0.780
1 (Positive)	0.897	0.875	0.886
Macro avg	0.830	0.838	0.833
Weighted avg	0.852	0.850	0.851

Overall:

- **Accuracy:** 0.850
- **Macro avg F1-score:** 0.833
- **Weighted avg F1-score:** 0.851