

Fitness-Based Recommender Systems for Reducing Sedentary Behaviour

Shogo Toyonaga

**A Thesis Submitted to the Faculty of Graduate Studies
In Partial Fulfillment of the Requirements
for the Degree of Masters of Science**

Graduate Program in Faculty of Electrical Engineering & Computer Science

**York University
Toronto, Ontario**

August 2025

© Shogo Toyonaga, 2025

Abstract

Obesity and sedentary behaviour represent one of the greatest global challenges to good health and wellbeing. The goal of the thesis is to promote physical activity among young adults by comparing the effectiveness of content-based and context-aware recommender systems on perceived post-intervention user experience, exercise motivation, and projected behaviour performance. Gender differences are explored. A 73-person user study compares recommender systems that solely focus on generating fitness plans (control group) against alternatives that incorporate psychosocial frameworks and explainability into the generation process (experimental group). The context-aware recommender systems provided the highest level of perceived post-intervention user experience, exercise motivation, and projected behaviour performance compared to the content-based recommender systems. Among females, the experimental group which leveraged persuasive design techniques showed numerical gains in exercise motivation and projected behaviour performance compared to the control group, however, the interaction effect was non-significant. Future work should investigate hybrid recommender systems in generating personalized exercise recommendations.

Acknowledgements

This Thesis would not have been remotely possible without the support and guidance from many colleagues, friends, and family members for whom I am very grateful for having in my life.

Firstly, I would like to thank my parents, John and Taeko Toyonaga for always having my back. Thank you for listening to me grumble about my work and for always pushing me to be a better person than I was yesterday. Your encouragement, empathy, and mentorship will always continue to be an invaluable source for growth.

Secondly, I would like to express my appreciation and gratitude to the friends that I made along the way: Abel Habte, Tony Tran, Jayant Varma, Oluwafemi Odu, Sarah Ejaz, Kingson Fung, Isaac (Oluwamayowa) Ibidun, Vrushank Vaghani, Mohammed Abdulmawjood, Cathy (Man Chau) Ng, Marco Antonio Lopez Arriaga, and Hooman Hoghooghi Esfahani. The personal and academic discussions that we shared during the very long days were invaluable in contributing to my research and personal development. I also appreciated the banter. A special shoutout is warranted for the 73 participants who took time out of their day to help me complete this work. I am truly grateful for their selflessness.

Thirdly, I am grateful for the opportunities to have worked alongside Dr. Barry Toyonaga, Dr. Celina Da Silva, Dr. Marios Fokaefs, and Ilo-Katryn Maimets. Thank you for deepening my knowledge about teaching pedagogy and scientific methodology. Thank you for allowing me to broaden my research area. It was truly a valuable and pleasant experience; I wouldn't have had it any other way.

Fourthly, I would like to extend my gratitude and appreciation for the Lassonde Graduate Program Assistants (Ciara Campbell, Huy Quang Doan, and Susanna Talanca) for helping me to navigate through the logistics of this degree. A shout out to the kind staff at the Lassonde Research Centre is warranted for their kindness and helpfulness in solving any research-related logistical problems.

Lastly, I would like to thank my supervisor, Dr. Kiemute Oyibo, and committee members, Dr. Shital Desai and Dr. Enamul Hoque Prince for their mentorship and guidance in learning to become a better scholar. This work would not be possible without their guidance, encouragement, and comprehensive feedback.

Contents

	Page
Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vii
List of Figures	ix
1 Introduction	1
1.1 Recommender Systems	1
1.2 Motivation	2
1.3 Research Objectives	3
1.4 Contributions	4
1.5 Thesis Structure	5
2 Literature Review	6
2.1 Explainable Recommendations	8
2.2 Trustworthy and Privacy-Preserving Recommendations	9
2.3 Applied Behavioural and Psychological Frameworks	11
2.4 Barriers to Fitness Adoption	12
2.5 Psychosocial Frameworks & Theories	16
2.6 Fogg Behavioural Model	16
2.7 Persuasive System Design Model	18
2.8 Social Cognitive Theory of Self-Regulation	19
2.9 COM-B Model	20
2.10 ComTech Model	21
2.11 Theory of Interactive Media Effects Model	22
3 Methodology	23
3.1 Design Science Research Methodology	23
3.1.1 Phase 1: Problem Discovery & Motivation	23
3.1.2 Phase 2: Formulation of Solution	24
3.1.3 Phase 3: Design & Development	26
3.1.4 Phase 4: Demo	27
3.1.5 Phase 5: Evaluation	32
3.1.6 Phase 6: Dissemination of Findings	37

3.2	Recommender Systems Development	38
3.2.1	Hardware	38
3.2.2	Content-based Recommender System	38
3.2.3	Context-Aware Recommender System	41
3.3	Explainability & User Experience	43
3.4	System Design	43
4	Results	45
4.1	User Study	45
4.1.1	Demographic Information	45
4.1.2	Pre-Intervention Motivation	47
4.1.3	Pre-Intervention Privacy Preferences	48
4.1.4	User Experience	50
4.1.5	Perceived Post-Intervention Motivation	55
4.1.6	Open-Ended Comments (Self-Efficacy)	58
4.1.7	Open-Ended Comments (Self-Regulation)	61
4.1.8	Projected Post-Intervention Behaviour Performance	62
4.1.9	Findings	65
4.1.10	Explainability (H1)	65
4.1.11	Psychosocial Frameworks (H2, H3)	65
4.1.12	User Experience (H4)	65
4.2	Computational Results	66
4.2.1	Content-based Recommender System	66
5	Discussion	67
6	Limitations and Future Work	73
6.1	Limitations in the Study Design	73
6.2	Limitations in System Design	74
6.3	Limitations Reported by Participants	75
6.4	Implications	78
7	Conclusion	80
	Bibliography	81
	Appendices	92
A	User Study (Instruments)	92
A.1	Consent Form	92
A.2	Pre-screening Questionnaire	94
A.3	Study Questionnaire	95

B	Datasets	102
C	Prompts	103
D	Application (Screenshots & Instruments)	105
E	Additional Results	107
E.1	Participant Demographics	107

List of Tables

1	Commonly Used Recommender Systems with their Corresponding Definitions	1
2	Commonly Studied Domains for Reduction of Sedentary Behaviour Using Fitness-Based Recommender Systems	6
3	Commonly Used Explainability Metrics Implemented in Fitness-Based Recommender Systems with their Definitions	8
4	Commonly Used Privacy-Preserving Interventions Used in Fitness-based Recommender Systems with their Definitions	9
5	Commonly Used Behavioural and Psychological Frameworks for Designing and Evaluating Fitness-Based Recommender Systems with their Definitions	11
6	Commonly Cited Barriers by Users to Increasing their Physical Activity levels	13
7	User Experience Design Techniques from the Adapted ComTech Model [1]	21
8	Explanation of How Entities and Constructs from Figure 8 Have Been Considered in System Design	24
9	Comparative Table of the Recommender Systems and Supported Features	32
10	Greene et al.'s Criteria for Conducting Mixed-Methods Research [2]	33
11	Applied Recruitment Strategies: Snowballing and Self-Selection Sampling	34
12	Razer Blade 2020 (Base): Hardware Specification of Laptop Used During Demo	35
13	Adapted Questionnaires Used to Evaluate the Fitness-Based Recommender Systems	36
14	Software Used for Statistical Analysis	37
15	Google Colaboratory: Hardware Specification	38
16	Murarka's Gym Exercise Dataset [3]	39
17	Context-Aware Recommender System: Required Features	41
18	Agent Configuration	43
19	Summary of Dependent Variables Studied from the User Study	45
20	Distribution of Participants by Gender and Physical Activity Classification into Groups and Systems	46
21	Aligned Rank Transform ANOVA Results based on Pre-Intervention Motivation	47
22	Pairwise Contrast Test for <i>Gender</i> \times <i>Group</i> \times <i>System</i> from Table 21. Contrasts with $p \geq 0.1$ have been filtered out for brevity.	48
23	Aligned Rank Transform ANOVA Results based on Pre-Intervention Privacy Attitudes	49
24	Pairwise Contrast Test for <i>Gender</i> \times <i>Group</i> \times <i>System</i> from Table 23. Contrasts with $p \geq 0.1$ have been filtered out for brevity.	50
25	Aligned Rank Transform ANOVA Results based on Perceived User Experience	52
26	Pairwise Contrast Test for the User Experience Factor from Table 25. Contrasts with $p \geq 0.1$ have been filtered out for brevity.	53
27	Pairwise Contrast Test for <i>System</i> \times <i>UX.Factor</i> from Table 25. Contrasts with $p \geq 0.1$ have been filtered out for brevity.	55

28	Aligned Rank Transform ANOVA Results based on Perceived Post-Intervention Exercise Motivation	57
29	Pairwise Contrast Test for <i>Gender × Group × Motivation.Factor</i> from Table 28. Contrasts with $p \geq 0.1$ have been filtered out for brevity. Abbreviations: SE Self-efficacy, SR Self-regulation	58
30	Chi-Square Test Results for Gender, Group, and System Variables Based on Sentiment Counts of Open-Ended Self-Efficacy Responses	59
31	Thematic Coding: Sample Barriers to Self-Efficacy Questions	59
32	Thematic Coding: Enablers on Open-Ended Exercise Self-Efficacy Questions	60
33	Chi-Square Test Results for Gender, Group, and System Variables Based on Sentiment Counts of Open-Ended Self-Regulation Responses	61
34	Thematic Coding: Strategies for Self-Regulating Exercise Adoption	61
35	Aligned Rank Transform ANOVA Results based on Projected Post-Intervention Behaviour Performance Assessment	64
36	Pairwise Contrast Test for <i>Gender × Group</i> from Table 35. Contrasts with $p \geq 0.1$ have been filtered out for brevity.	64
37	Weighted F1-Score on Training, Testing, and Validation Splits for the Content-Based Recommender System	66
38	Classification Report Results on Testing Split for the Content-Based Recommender System	66
39	Subset of Open-ended feedback on the usability of the recommender system(s)	67
40	Subset of Open-ended feedback on the Context-Aware Recommender System (Separated by Study Group)	70
41	Identified Themes from Open-Ended Participant Feedback for Future Improvements to the Content-Based Recommender System	75
42	Identified Themes from Open-Ended Participant Feedback for Future Improvements to the Context-Aware Recommender System	76
A.2.1	Get Active Questionnaire [4]	94
A.2.2	Cognitive Change Questionnaire [5]	95
A.3.1	Physical Activity Enjoyment Scale (PACES-8) Questionnaire [6]	96
A.3.2	Privacy Attitudes Questionnaire (PAQ) [7]	96
A.3.3	Adapted ComTech (HUED-CUP) Questionnaire [1]	97
A.3.4	Self-Efficacy for Exercise Scale (SEE) [8]	98
A.3.5	Adapted Questionnaire from the Exercise Goal-Setting Scale (EGS) and Exercise Planning and Scheduling Scale (EPS) [9]	99
E.1	Participant Demographics: Age Groups	107
E.2	Participant Demographics: Ethnicity	107
E.3	Participant Demographics: Education Level and Major of Study	108

List of Figures

1	Fogg Behavioural Model (FBM) [10]	17
2	Persuasive System Design (PSD) Model [11]	18
3	Social Cognitive Theory of Self-Regulation (SCT) [12, 13]	19
4	Capabilities, Opportunities, and Motivation (COM-B) [14, 15]	20
5	Adapted HUED-CUP Framework [1]	21
6	Theory of Interactive Media Effects (HAI-TIME) Model [16]: Cues and Action Routes	22
7	Design Science Research Methodology (DSRM) Process Model [17]: Problem Discovery & Motivation, Formulation of Solution, Design & Development, Demo, Evaluation, Dissemination of Findings	23
8	Applied Framework for Projected Behaviour Performance Adapted from [1, 2, 18, 19]. Elements of the Framework are Leveraged from FBM [10], PSD [11], SCT [12], COM-B [14], HUED-CUP [1], and HAI-TIME [16]	24
9	Keyson and Parson's Adapted Iterative Design Process [20] Adapted from Willege's Three Stage Iterative Design Process [21]	26
10	User interfaces for the support recommender systems	27
11	Relevance of Exercise Recommendations from the Experimental Content-Based Recommender System	28
12	Explainable Feature Relevance Plots from the Experimental Content-Based Recommender System	28
13	Exercise Plan Generated by the Experimental Content-Based Recommender System	29
14	Persuasive Techniques to Increase Exercise Motivation in the Experimental Content-Based Recommender System	29
15	Exercise Plan Generated by the Experimental Context-Aware Recommender System	30
16	Diet Plan Generated by the Experimental Context-Aware Recommender System	30
17	Relevance of Exercise Recommendations from the Control Content-Based Recommender System	30
18	Exercise Plan Generated by the Control Context-Aware Recommender System	31
19	Diet Plan Generated by the Control Context-Aware Recommender System	31
20	Content-based Filtering Recommender System: Training and Evaluation Pipeline	40
21	Context-Aware Recommender System: Agent Workflow	42
22	Mean exercise motivation scores among the study groups and recommender systems. Error bars represent $\pm 1 SD$	47
23	Mean exercise motivation scores among the <i>Gender</i> \times <i>Group</i> \times <i>System</i> interactions. Error bars represent $\pm 1 SD$	48
24	Mean privacy attitude scores among the study groups and recommender systems. Error bars represent $\pm 1 SD$	49
25	Mean privacy attitude scores among the <i>Gender</i> \times <i>Group</i> \times <i>System</i> interactions. Error bars represent $\pm 1 SD$	50

26	Mean User Experience scores among the study groups and recommender systems. Error bars represent $\pm 1 SD$	51
27	Mean User Experience scores among the <i>Gender</i> \times <i>Group</i> \times <i>System</i> interactions. Error bars represent $\pm 1 SD$	51
28	Mean User Experience scores based on the varying ComTech Factors. Error bars represent $\pm 1 SD$	53
29	Mean User Experience scores based on the <i>System</i> \times <i>UX.Factor</i> interactions. Error bars represent $\pm 1 SD$	54
30	Mean User Experience Scores based on Subgroups: Gender, System, and UX.Factor. Error bars represent $\pm 1 SD$	54
31	Mean Post-Intervention Exercise Motivation scores among the study groups and recommender systems. Error bars represent $\pm 1 SD$	56
32	Mean Post-Intervention Exercise Motivation scores based on the <i>Gender</i> \times <i>Group</i> \times <i>System</i> interactions. Error bars represent $\pm 1 SD$	56
33	Mean Post-Intervention Exercise scores among the study groups and recommender systems. Error bars represent $\pm 1 SD$	63
34	Mean Post-Intervention Exercise scores based on the <i>Gender</i> \times <i>Group</i> \times <i>System</i> interactions. Error bars represent $\pm 1 SD$	63
35	Proposed High-Level Multi-Agent System for Generating Personalized Fitness Plans. Nodes Represent Individual Agents and Edges Represent Communication Between Different Actors/Critics.	79
B.1	Murarka’s Dataset [3]: Visualization After Removing Labels with less than 4 Occurrences	102
C.1	Prompt Template: Context-Aware Recommender System	103
C.2	Control Prompt: Context-Aware Recommender System	103
C.3	Experimental Prompt: Context-Aware Recommender System	103
C.4	Experimental Prompt: Content-based Recommender System	104
D.1	Control Group: Introduction Page (Privacy Policy)	105
D.2	Experimental Group: Introduction Page (Privacy Policy)	105
D.3	Experimental Group: Verbose Mode of Context-Aware Agent	106

1 Introduction

1.1 Recommender Systems

Recommender systems play a critical role in personalizing user interactions and experiences in domains such as e-commerce (e.g., literature, music) and personal health records [22, 23, 24]. Their ubiquity in modern society can be explained by rapid advancements in mobile technology and sensors, enabling a wealth of autonomous data collection readily available for machine learning applications [23, 25].

Recommender systems are classified and implemented differently based on the required features to generate the corresponding label. Table 1 summarizes the different types of recommender systems based on their definitions and feature characteristics that have been used across various domains. Applications of recommender systems within the fitness domain remain an understudied area, thereby creating many opportunities for future work [26, 27]. For example, a systematic review by Toyonaga and Oyibo [28] that evaluates fitness-based recommender systems identified that 40.74% of studies did not use human participants and only relied on computation-based metrics (e.g., predictive accuracy, F1-score). They also identified that 31% of the evaluated studies did not leverage any theoretical frameworks to design or evaluate the systems; rather, the system design solely relied on ad hoc principles. Lastly, recommendation explainability (28%) and privacy-preserving methods (13%) were shown to be lacking in the systems design and implementation.

Table 1: Commonly Used Recommender Systems with their Corresponding Definitions

Recommender System	Definition
Collaborative Filtering	Using profiles of similar users (e.g., preferences) to determine relevant products or services of interest [22].
Content-based Filtering	Using information about products or services (e.g., cost, descriptions, themes) to determine relevant suggestions to a user [22].
Context-Aware	Using user and/or content-based information in addition to dynamic contextual cues that affect user preferences. Examples include the time of day, location, or weather [29].
Demographic	Using personal information and biomarkers (e.g., age, gender, ethnicity, weight, height) to determine a relevant suggestion [29].
Hybrid	Using a combination of two or more types of recommender system architectures.
Knowledge-based Reasoning	Using a database of information to support rule-based reasoning (e.g., implications) in the generation of relevant recommendations [30].

1.2 Motivation

Current fitness-based interventions that aim to reduce sedentary behaviour (e.g., hiring a personal trainer) have been shown to be ineffective due to monetary and accessibility-related barriers [31, 25]. Furthermore, the one-size fits all approach for designing interventions for a particular individual without consideration of dynamic preferences and capabilities has resulted in low user engagement and retention rates [23, 32, 33]. For example, a study showed that 50% of participants from the United States who signed up for a fitness class withdrew within the first six months [32].

The lack of effective fitness-based interventions which reduce sedentary behaviour is problematic as lack of physical activity among the global population has presented itself as one of the greatest threats to good health and well-being [34]. In 2020, the World Health Organization (WHO) announced that the global adult (28%) and child population (81%) were not sufficiently physically active, citing an obesity epidemic [34, 31, 29].

Chronic obesity significantly increases risk factors related to cardiovascular, pulmonary, cognitive (e.g., Alzheimers), and diabetes-related illnesses, which significantly reduce one's quality of life [34, 31, 35, 36, 37]. Obesity prevention is especially pertinent for children as the risk factors remain even if they lose the excess weight later on in life [34]. At a global level, chronic obesity results in the loss of billions of dollars through healthcare expenditure (e.g., medicine, treatment) and an inability to work due to presenteeism, absenteeism, and premature mortality [34].

Certain populations such as women, university students, people living on low-income, and residential school survivors are at heightened risk of obesity [36, 38, 39]. They require specialized consideration in the development of these intervention systems. While their particular cause of obesity can be explained due to biological differences, work-life balance, inaccessibility to relevant services, or affordability of healthy foods, these factors demonstrate the necessity in developing theory-driven context-aware recommender systems in the fitness domain [36, 39].

This thesis seeks to develop a personalizable fitness-based recommender system that leverages theory-driven behavioural and psychological frameworks. The experimental system will be evaluated against a control system that does not incorporate factors of explainability and user trust. The target demographic represents young adults, aged 18-30 years old with no physical or cognitive disabilities that can be readily determined.

1.3 Research Objectives

This thesis aims to address the following research questions:

RQ1: How does explainability affect the perception and adherence to fitness-based recommendations?

To investigate how explainability affected user perception, we provided participants with SHapely Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanation (LIME) bar plots for each prediction. Furthermore, we communicated the probability estimates for the top 5 most relevant recommendations. Lastly, for each prediction, we used a Large Language Model (LLM) with Chain-of-Thought (CoT) prompting to explain why the recommendation was made from the participant’s user profile and how to integrate it into their daily routines using techniques from persuasive design.

RQ2: How does the application of theoretical frameworks grounded in social science and psychology used to design fitness-based recommender system affect the perceived utility of the fitness plans?

To investigate the effectiveness of theoretical social and psychology-based frameworks in the design and evaluation of fitness-based recommender systems, we compared the results of a control and experimental group using Aligned Rank Transform ANOVA [40] and post-hoc contrast testing (ART-C) [41]. Thematic coding, sentiment analysis, and text mining of open-ended questionnaire data was used to evaluate the qualitative aspects of the study results. The control group implemented the fitness-based recommender systems without persuasive design or behaviour-based considerations. Furthermore, explainability was not supported through logical text-based reasoning (e.g., Chain-of-Thought Prompting) and data visualizations (e.g., SHAP, LIME).

RQ3: How does user experience (e.g., trust, privacy, and perception) affect user attitudes and adherence to fitness-based recommender systems?

Moreover, to investigate user trust, we implemented Sundar’s Human-AI Interaction Model based on the Theory of Interactive Media Effects (HAI-TIME) [16] into the system. The cue-based route leveraged Bandwagon, Social, and Authority-based heuristics in the LLM prompts and design of each recommender system.

To investigate how privacy affected user attitudes, we minimized data collection as much as possible and did not save the provided features to each of the recommender systems. Furthermore, we use open-source models in the experimental condition (mistralai/Mistral-Nemo-Instruct-2407) and closed-source models (openai/gpt-4o) in the control condition. Lastly, the privacy policies on the main screen of the experimental intervention were written differently to immediately reflect these changes to the participant prior to interacting with the recommender systems.

1.4 Contributions

Completion of this thesis has resulted in the following contributions:

- **Contribution 1:** We conducted a comprehensive systematic review [28] on fitness-based recommender systems, which uncovered unmet needs from a hybrid technological and social science perspective [28]. The manuscript outlined future directions for fitness-recommender systems and provided a framework for implementing explainable, privacy-preserving recommender systems which shaped the development of this work.
- **Contribution 2:** We developed an explainable, privacy-preserving recommender system that leveraged social science, psychology-based, and persuasive technology design frameworks and evaluated it in a user study. The Design Science Research Methodology (DSRM) Process Model [17] was used to guide this research.
- **Contribution 3:** We used intelligent agents with search and retrieval-based tools (e.g., Retrieval Augmented Generation) to generate personalized fitness plans for participants. The plans could be adapted and re-designed according to changing preferences, physical capabilities, or design frameworks in real-time, maximizing system trust.
- **Contribution 4:** We leveraged multiple theoretical frameworks such as the Fogg Behavioural Model (FBM) [10] and Persuasive System Design Model (PSD) [11] to personalize fitness recommendations. We focused on maximizing user adherence to the recommendations by emphasizing motivation, self-efficacy, and actionable triggers.
- **Contribution 5:** We implemented several recommender systems based on varying levels of required information (e.g., exercises, physical condition, personal preferences) to evaluate how privacy affects user attitudes and behaviour towards exercise.

The contributions of this work are also in line with five of the 17 United Nations Sustainable Development Goals (UNSDGs) as follows:

1. **Good Health & Well-being (SDG3):** Long-term exercise adoption increases a persons quality of life by reducing risk factors related to chronic physical and cognitive diseases such as type 1 or 2 diabetes, hypertension, and Alzheimer's [34, 31, 35, 36, 37].
2. **Quality Education (SDG4):** Programs that teach students about healthy behaviours related to obesity prevention and weight management (e.g., exercises & diets) are lacking or inaccessible to the broad public [39]. Fitness-based recommender systems can be used as a cost-effective and readily-accessible resource that drives University programs aiming to address weight management in a personalized and empathetic manner.

3. **Decent Work & Economic Growth (SDG8):** Obesity and its related comorbid diseases result in significant declines in workforce productivity due to presenteeism, absenteeism, and premature death [34]. Furthermore, the increase in healthcare services caused by obesity results in billions of dollars in healthcare expenditure through medical care and prescriptions. This can be minimized through the use of personalized fitness-based recommender systems.
4. **Reduced Inequalities (SDG10):** Existing interventions which seek to reduce obesity among the general population often fail due to a one-size-fits-all approach [27]. A lack of consideration towards barriers such as scheduling, injuries, available resources, or social influences (e.g., gender [42]) are solved through the application of intelligent agents with proper tools.
5. **Partnerships for the Goals (SDG17):** Fitness-based recommender systems are understudied and lack a commonly accepted framework for their design and evaluation [27, 26]. This work provides a framework for the design and evaluation of fitness-based recommender systems grounded in social science and psychology literature. Furthermore, common barriers resulting from existing recommender systems such as a lack of explainability, trust, and user privacy are addressed to increase long-term user engagement and retention.

1.5 Thesis Structure

Chapter 1 introduces the concept of recommender systems and outlines the contributions of this work. Chapter 2 summarizes the theory-driven frameworks used in this work and discusses how recommender systems have been administered in the fitness domain. The literature review identifies key areas for future work. Chapter 3 outlines the methodology used in addressing our research questions. Chapter 4 outlines the results of the experiments. Chapter 5 discusses and contextualizes the results with respect to previous work. Chapter 6 identifies implications and future work. Lastly, Chapter 7 summarizes the undertaken work and reiterates the fundamental contributions.

2 Literature Review

This chapter outlines previous work that provides context for the contribution of this thesis. Furthermore, it highlights and outlines behaviour and psychology-based frameworks (Chapter 2.5) that will be used to support the research design. These frameworks have been chosen to personalize fitness plans by simultaneously considering motivation, technology acceptance, and environmental factors.

Toyonaga and Oyibo’s [28] systematic review on fitness-based recommender systems outlined active areas in which research has been undertaken. Table 2 summarizes each of the studied domains with relevant references.

Table 2: Commonly Studied Domains for Reduction of Sedentary Behaviour Using Fitness-Based Recommender Systems

Target Behaviour	Activities	References
Reduction of Sedentary Behaviour	Workout Correction	Jaiswal et al. [43]
	Diets	Mahyari and Pirolli [44]
	User Dropout	Kulev et al. [45]
	Activity Classification	Coppens et al. [26]
	Walking Routes	Coppens et al. [46]
		Coppens et al. [47]
		Tragos et al. [48]
		Ono et al. [49]
		Lin et al. [50]
		Li et al. [51]
		Sengan et al. [52]
		Dharia et al. [53]
		Mahyari et al. [54]
Ali et al. [30]		
Alcaarez-Herrera and Palomares [55]		
Chowdhury et al. [56]		
Pilloni et al. [57]		
Kadri et al. [58]		
Takama et al. [59]		
Minakata et al. [60]		
Active Ageing (Elderly Care)	Exercises	Martinez-Martin et al. [31]
	Diets	Palumbo et al. [61]
		Wongpun and Guha [62]
Strength Training	Exercise Guidance	Tran et al. [63]

Table 2 (continued)

Target Behaviour	Activities	References	
Weight Loss	Exercise Guidance	Liu et al. [25]	
	Exercises	Zhao et al. [64]	
	Diets	Ichino et al. [65]	
	Fitness Plans		Agans et al. [27]
			Zhao et al. [66]
			Afzal et al. [67]
			Yang and Zhao [68]
			Ali et al. [69]
			Basnayake et al. [70]
			Palomares et al. [71]
			Gomez-del-Rio et al. [72]
			Alcaarez-Herrera et al. [73]
			Sun et al. [74]
Balpande et al. [75]			
Gasparetti et al. [76]			
Marathon Training	Training Plans	Berndsen et al. [77]	
	Diets	Berndsen et al. [78]	
		Feely et al. [79]	
		Donciu et al. [80]	
Management of Chronic Illness	Exercises	Ferretto et al. [81]	
	Fitness Plans	Cavallo and Toumazou [82]	
	Diets	Vairale and Shukla [83]	
		del Rio et al. [35]	
		Alian et al. [84]	
Xie and Wang [85]			
Rehabilitation (Therapy)	Exercises	Chang et al. [86]	
	Diets	Gmez-Portes et al. [87]	
	Fitness Plans	Gonzalez-Gonzalez et al. [88]	
	Reminders (Motivation)	Ishraque et al. [89]	
		Blasco et al. [90]	

2.1 Explainable Recommendations

Research into the explainability of fitness-based recommender systems, a type of information system that generates exercise and dietary plans based on user data, has been scant [27, 26, 73, 71, 46, 79, 54]. Table 3 outlines identified techniques and studies which leveraged explainability into the recommendation process.

Table 3: Commonly Used Explainability Metrics Implemented in Fitness-Based Recommender Systems with their Definitions

Implementation	Definition	References
Rule-based Reasoning	Using logical implications to generate relationships between entities from knowledge-bases to explain an algorithms decisions.	Ali et al. [30] Wongpun and Guha [62] Palomares et al. [71] Lin et al. [50] Afzal et al. [67] Gonzalez-Gonzalez et al. [88] Gmez-Portes et al. [87] Alian et al. [84]
Feature Importance	Using statistical analyses and properties of machine learning models such as impurity to determine the impact of features on the prediction of a particular label.	Pilloni et al.[57] Zhao et al. [64] Yang and Zhao [68] Feely et al. [79] Coppens et al. [47]
Text-based Reasoning	Using textual argumentation or explanations to help support adherence to the recommended exercise or fitness plan.	Feely et al. [79] del Rio et al. [35]
Risk Visualizations	Visualizing risky locations for participant safety on a 3-dimensional map interface.	Minakata et al. [60]

Of the studies which leveraged rule-based reasoning, 75% (6 / 8) were user or hybrid studies that evaluated recommendation quality through metrics such as accuracy or agreement of relevance from target users and healthcare experts [62, 71, 50, 67, 88, 84]. None of the studies evaluated whether the explainable rules or the data sources by themselves were influential in affecting user trust in the system. For example, evaluating whether the rules were easily comprehensible to the users was not evaluated.

Similarly, of the user or hybrid studies which calculated and visualized feature relevance (3/5, 60%), none of them evaluated whether the visualizations or descriptions could affect users’ perception or belief in the machine learning model. They were only used to describe the model quantitatively within the article [57, 64, 68, 79, 47].

Lastly, in the case of user or hybrid studies which used text-based reasoning or risk visualizations, only one user study (1/3, 33%) qualitatively evaluated the impact of explainable recommendations [79, 35, 60]. del Rio et al. [35] demonstrated that using fact-based reasoning in their fitness-based recommendations resulted in participants having a positive perception towards the intervention.

2.2 Trustworthy and Privacy-Preserving Recommendations

Table 4 outlines the current implementations of trust and privacy-preserving methods in fitness-based recommender systems from Toyonaga and Oyibo’s systematic review [28].

Table 4: Commonly Used Privacy-Preserving Interventions Used in Fitness-based Recommender Systems with their Definitions

Implementation	Definition	References
User Authentication	Personal data is safeguarded against malicious actors by requiring user authentication (e.g., username and passwords) [91].	Afzal et al. [67] Banos et al. [91]
Encryption	Using an algorithm to obfuscate the original data (plaintext), making it uninterpretable (ciphertext) without the proper key. Often, the key is related to a user identity metric [52].	Ishraque et al. [89] Sengan et al. [52]
European GDPR Compliance	Governmental regulations that protect user privacy through enforcement’s such as user anonymization and clear privacy policies. Often, these policies outline what data is collected, how it used, and how to opt out of the policy. [68].	Yang and Zhao [68] del Rio et al. [35]
Anonymization or Prediction of Personal Information	Anonymization of data [48] refers to the removal of information from participant data which could reveal their personal identity. Prediction of personal data refers to using models trained on non-personal information to predict biomarkers such as gender or culture [25].	Tragos et al. [48] Liu et al. [25]

Of the studies which implemented privacy-preserving policies, 50% (4/8) were user studies that evaluated metrics such as usability, user experience, and opinions on recommendation relevance [67, 68, 35, 48]. However, it was not made clear if participants were made aware of the privacy policies and whether this impacted their perception of the systems quality and trust-worthiness.

Sun et al. [74] evaluated how features of recommender systems such as the inclusion or exclusion of social media data and affordance of user control affect health-based outcomes. They leveraged the Theory of Interactive Media Effects (HAI-TIME) model to design and evaluate the study [16]. It was found that of the 6 personalization approaches, those which leveraged social media data in the generation of recommendations were perceived as least favourable. Additionally, while affording users with control in the recommendation process received positive results, the effect was only statistically significant in the group which initially invoked threat to user identity. Questionnaire data demonstrated that the majority of negative sentiment around the social media conditions involved (1) a lack of relevant information, (2) privacy concerns, and (3) uncertainty of the intervention's abilities.

While the study did utilize the cue (heuristics) and action-based (affordance) routes with success through increasing user trust in the system, persuasive design frameworks were not utilized. Furthermore, the recommender systems were not implemented; a Wizard of Oz protocol was used as the fitness recommendation was predetermined and static for every condition. Lastly, only 3 heuristics were implemented into the prototype; bandwagon, identity, and machine. Future work can evaluate how other heuristics affect user trust in the system in addition to implementing other recommender systems (e.g., context-aware, hybrid).

Liu et al. [25] developed and evaluated a privacy-friendly recommender system that supported personalized distance, speed, and heart rate (zone) thresholds. Explicit data collection of personal information (e.g., age, gender, height) was circumvented through the use of Tensor Decomposition and sensor-based datasets to predict user information. Once all of the relevant features were obtained, a Multilayer Perceptron (MLP) and Bi-directional Long-Short Term model were trained and evaluated to support fitness guidance. Evaluation metrics (Root Mean Squared Error, Mean Absolute Error) demonstrated that the model was capable of making accurate recommendations in spite of using data augmentation techniques.

The study outcomes were limited as only computation-based metrics were evaluated. For example, the perception of the privacy-preserving technique could not be evaluated. This is significant as users may prefer to select implicit data entry rather than having their personal information predicted. By not giving users control in the recommendation process, trust in the system may be reduced, according to the HAI-TIME framework [16].

Sengan et al. [52] implemented a secure cloud-based recommender system that leveraged personal information and sensor-based data to suggest daily workouts. Furthermore, their architecture supported homomorphic encryption standards on user data to safeguard against privacy violations. Computation-based evaluation (Mean Absolute Error, Root Mean Square Error) on three testing datasets demonstrated that the recommender system was effective and robust against attack vectors such as bad-mouth or on-off attacks.

Similar to Liu et al. [25], by not leveraging human participants, it remains unclear how users would perceive the effectiveness of the privacy-preserving intervention. For a non-technical person, encryption may seem insufficient in preserving personal information without trust in the Organization.

2.3 Applied Behavioural and Psychological Frameworks

Table 5 outlines theoretical frameworks which have been used to design or evaluate fitness-based recommender systems in user and computational-based studies.

Table 5: Commonly Used Behavioural and Psychological Frameworks for Designing and Evaluating Fitness-Based Recommender Systems with their Definitions

Theoretical Framework	Definition	References
Behaviour Change Theory & Self-Efficacy	Developing interventions through planned activities which seek to change pre-existing behaviours [14]. Often, this is done through considering factors such as motivations, capabilities, and self-efficacy (e.g., Belief [18]).	Li et al. [51] Mahyari and Pirolli [44] Palumbo et al. [61] Cavallo and Toumazou [82] Ichino et al. [65] Coppens et al. [47] Coppens et al. [26]
Gamification	The personalization of an intervention through the consideration of factors such as user modeling, dynamic elements, and rewards (e.g., points). Often, spaced-repetition is emphasized to help promote learning objectives. [66, 92].	Gmez-Portes et al. [87] Zhao et al. [66] Gomez-de-Rio et al. [72] Zhao et al. [64] Gonzalez-Gonzalez et al. [88]
User Modelling & Self-Reflection	Integration of personalized recommendations to users based on feedback [46]. Often, it is obtained or inferred through the consideration and contextualization of factors relating to the current user’s status, profile, and weather conditions. Self-Reflection refers to an individual’s comparison of abilities against themselves (personal) or others (social) to motivate behaviour change [45].	Ishraque et al. [89] Kulev et al. [45] Coppens et al. [46] Sun et al. [74] del Rio et al. [35]

Mahyari and Pirolli [44] implemented a mobile-based collaborative filtering recommender system that generated personalized exercises and health goals. It incorporated Behaviour Change Theory and Self-Efficacy into the design of the intervention. The interconnected recurrent neural network was trained to predict exercise recommendations based on a high likelihood of user adherence and exercise completion. While the computational-based study had a positive outcome with high accuracy, there were limitations in study generalizability and data sparsity. Firstly, not evaluating the intervention with participants meant that relevant outcome metrics such as user behaviour and psychological states could not be analyzed. Secondly,

the model suffered from data sparsity issues; accuracy diminished for underrepresented classes that diverted from a one-size fits all approach. As user engagement drops, the model is unable to improve or maintain satisfactory recommendation relevance given its current implementation.

Gomez-del-Rio et al. [72] developed a web-based hybrid recommender system for obese children targeting relevant dietary and physical activity suggestions. It was integrated into an exergame and designed using the Quantified-Self Model [93]. The Quantified-Self Model was integrated with Gamification techniques (e.g., rewards) in order to increase personalization and motivational factors by delivering relevant real-time health metrics to the user. Additionally, a trigger was built into the system using a multi-armed bandit approach to determine the best times to notify a user for increasing their physical activities. While the randomized control study outcome with 46 participants was mostly positive, questionnaire data revealed that any weight lost during the study was re-gained. However, the participants in the experimental group reported significantly higher self-efficacy and positive emotions towards exercises. Additionally, their knowledge of healthy behaviours (diet and exercise-based) was significantly improved after completing the study. A limitation of the study was that the user interface and usability of the exergame was not evaluated. This may have subsequently affected the effectiveness of the exergame. Furthermore, small group sizes may have led to noise in the statistical testing and analysis.

Sun et al. [74] performed a user study to evaluate how fitness-based recommender systems are perceived when different implementations and privacy policies are used. Content-based and collaborative-filtering recommender systems were prototyped, designed, and evaluated using the Human-AI Interaction Model based on the Theory of Interactive Media Effects (HAI-TIME) model [74, 16]. In each of the conditions, the provided fitness recommendation was static to avoid introducing biases and noise into the results. The HAI-TIME model states that a user's trust in the system is increased through cue-based (heuristic) or action routes [16]. Heuristics increase a user's trust in the system by minimizing mental resources required to evaluate credibility (e.g., BandWagon or Authority-based endorsements). The action-based route increases user trust by enabling multiple pathways for how recommendations are generated. In this way, the user can select an intervention which closely aligns with their personal and/or privacy preferences. While the study outcome was mostly positive and proved that cue and action routes increase user trust, there were generalizability limitations insofar as other recommender system architectures were not considered (e.g., hybrid, context-aware). Furthermore, the study could have been extended by considering the use of persuasive message design frameworks [94] in the representation of exercise suggestions.

2.4 Barriers to Fitness Adoption

Table 6 outlines barriers to increasing physical activity which have been addressed through the implementation of fitness-based recommender systems.

Table 6: Commonly Cited Barriers by Users to Increasing their Physical Activity levels

Barrier	Definition	Reference
Healthcare Accessibility	Limited resources to adequately provide patients with access to healthcare professionals or services such as elderly care [62] or therapy specialists [87]. Professionals are required to create personalized, effective, and actionable health plans. Additionally, healthcare accessibility can be related to barriers such as language, education level, location, and monetary costs which prevent individuals from seeking help [84].	Wongpun and Guha [62] Ishraque et al. [89] Gmez-Portes et al. [87] Alian et al. [84]
Personalization	Acquiring recommendations with methodologies that closely align with the users preferences and capabilities [80]. Furthermore, effective recommendation personalization must balance accuracy, diversity, and privacy preferences to reduce the likelihood of user dropout [48, 26, 74].	Donciu et al. [80] Tragos et al. [48] Mahyari and Pirolli [44] Blasco et al. [90] Balpande et al. [75] Palomares et al. [71] Vairale and Shukla [83] Alcaez-Herrera et al. [73] Gasparetti et al. [76] Feely et al. [79] Berndsen et al. [78] Chang et al. [86] Alcaez-Herrera and Palomares [55] Ali et al. [69] Ali et al. [30] Afzal et al. [67]
Awareness of Medical Condition	Creating interventions which are capable of classifying and communicating the state of a user's medical condition or sedentary behaviours (e.g., Kadri et al. [58]) effectively while providing opportunities for healthier living [81].	Ferretto et al. [81] Kadri et al. [58] Basnayake et al. [70]

Table 6 (continued)

Barrier	Definition	Reference
Privacy Policies	Implementing measures to protect a user's personal information. Table 4 outlines common privacy policies which have been implemented in existing studies with their definitions.	Sengan et al. [52] Sun et al. [74] Liu et al. [25]
Dislike or Fear of Injury	Addressing negative sentiment towards exercise-related behaviours through focusing on barriers, social components, explainability, and self-efficacy [27, 60].	Agans et al. [27] Xie and Wang [85] Jaiswal et al. [43] Minakata et al. [60]
Scheduling Conflicts	Generating exercise recommendations and notifying the user at the optimal time to maximize the probability of adherence [49]. Often, this is accomplished through incorporating a user's calendar availability or other related contextual factors such as weather into the recommendation engine.	Ono et al. [49] Chowdhury et al. [56] Martinez-Martin et al. [31] Li et al. [51]
Self-Efficacy	Increasing a user's intrinsic motivation and belief in their capabilities to complete exercises and change behaviour. [60].	Mahyari et al. [54] Ichino et al. [65] del Rio et al. [35] Cavallo and Toumazou [82] Lin et al. [50] Coppens et al. [26]

Table 6 (*continued*)

Barrier	Definition	Reference
Motivation	Personalizing recommendations to keep users interested in completing follow-up exercises. For example, Palumbo et al. [61] addressed motivational problems by recording user-reported exercise experiences and maximizing pleasant activities. Gamification could also be used by leveraging motivational factors such as rewards to keep users interested in the intervention [66, 64].	Palumbo et al. [61] Zhao et al. [66] Gomez-del-Rio et al. [72] Zhao et al. [64] Gonzalez-Gonzalez et al. [88] Tran et al. [63] Dharia et al. [53] Pilloni et al. [57] Yang and Zhao [68] Kulev et al. [45] Berndsen et al. [77] Coppens et al. [47] Coppens et al. [46] Takama et al. [59]

Wongpun and Guha [62] implemented and evaluated a web-based hybrid recommender system to assist informal caregivers with elderly care. Given a user’s personal information and daily activities, the case-based reasoning engine was able to produce relevant dietary, exercise-based, and emotional-support suggestions. Three healthcare experts were used to assess the quality of the generated recommendations based on 5 case studies, achieving a 78.33% appropriateness. While the study outcome was positive, generalizability was reduced through a low population size. Furthermore, the intervention was built using a small dataset of 51 elderly participants. Data augmentation of the preliminary recommendations was not considered to strengthen the robustness of the system.

Alcaez-Herrera and Palomares [55] implemented and evaluated a context-aware recommender system to help support the generation of exercise and diet pair recommendations. Unlike previous studies which only consider recommendations in isolation to a trained dataset or database, the intervention used an evolutionary genetic algorithm to optimize personalization. After obtaining the user’s personal information, preferences, and constraints, the algorithm would iterate and mutate several times until the fitness function achieved a sufficient convergence. The results of the user study among 54 participants was mostly positive with the questionnaires demonstrating high user satisfaction and recommendation relevance. However, the intervention was also shown to make incorrect recommendations when the participant’s primary goal was to build muscle. This demonstrates a limitation of the system, requiring the aid and collaboration of domain experts.

Chowdhury et al. [56] implemented and evaluated a desktop-based context-aware recommender system for generating fitness advice. Interaction was facilitated through a chatbot user interface and a deep neural network architecture. Training loss with respect to the Recipes 1M and NutriStrategy datasets with anonymized student calendar availability demonstrated positive results, however, testing loss was not reported. The study outcome was limited as users did not interact with the system.

Similarly, Blasco et al. [90] implemented a mobile context-aware recommender system that aimed to address low motivation and self-efficacy through personalization. The intervention used a chatbot interface to provide rehabilitation plans for post-operative surgery using reminders, exercises, and explanations. Questionnaire data from participants of the study demonstrated mixed results; there were statistically significant effects in terms of quality of life, pain, and adherence to the rehabilitation plan, however, shoulder strength and mobility remained neutral. The study was limited by a small study duration, sample size, and effect size. Lastly, future work could aim to integrate the intervention through large language models or intelligent agents.

2.5 Psychosocial Frameworks & Theories

This section outlines each of the psychosocial frameworks and theories that were used to help facilitate the design and evaluation of theory-driven fitness-based recommender systems. Each of the frameworks emphasize important constructs such as motivation, environment, technological acceptance, privacy-preferences, and user autonomy, however, individually, none of them are adequate in accounting for all of the constructs simultaneously. As a result, the models have been joined together (See Figure 8) to propose a more holistic and all-encompassing framework.

2.6 Fogg Behavioural Model

The Fogg Behavioural Model (FBM) [10] (Figure 1) is a psychology-based theoretical framework that guides the analysis, implementation, and evaluation of persuasive technologies. It consists of three factors; motivation, ability, and triggers, which can facilitate the performance of a desired behaviour. Assuming that all of the three premises are present and sufficiently high, the desired behaviour is more likely to occur as the activation threshold is met.

Firstly, motivation refers to the inherent will to perform an action. It is often evaluated based on emotions or reactions that the desired behaviour elucidates; pleasure, pain, hope, fear, social acceptance, and social rejection. Each of these metrics are nominal; it is up to the designer to consider behavioural reactions and integrate what is necessary.

Secondly, ability refers to the difficulty or effort required to perform a behaviour. FBM posits that six requirements must be met to support ability through simplicity; time to complete the action, monetary constraints, physical capabilities, cognitive capacity, social norms, and routine integration. Similar to motivation, the designer must consider each of the six requirements and design an intervention for a specific target audience as abilities are dynamic.

Lastly, triggers refer to a call to action; the ways in which a user is notified to perform a behaviour. There are three types of triggers based on sparks, facilitators, or signals. Each trigger is designed to help overcome low motivation or lack of, real, or imagined abilities.

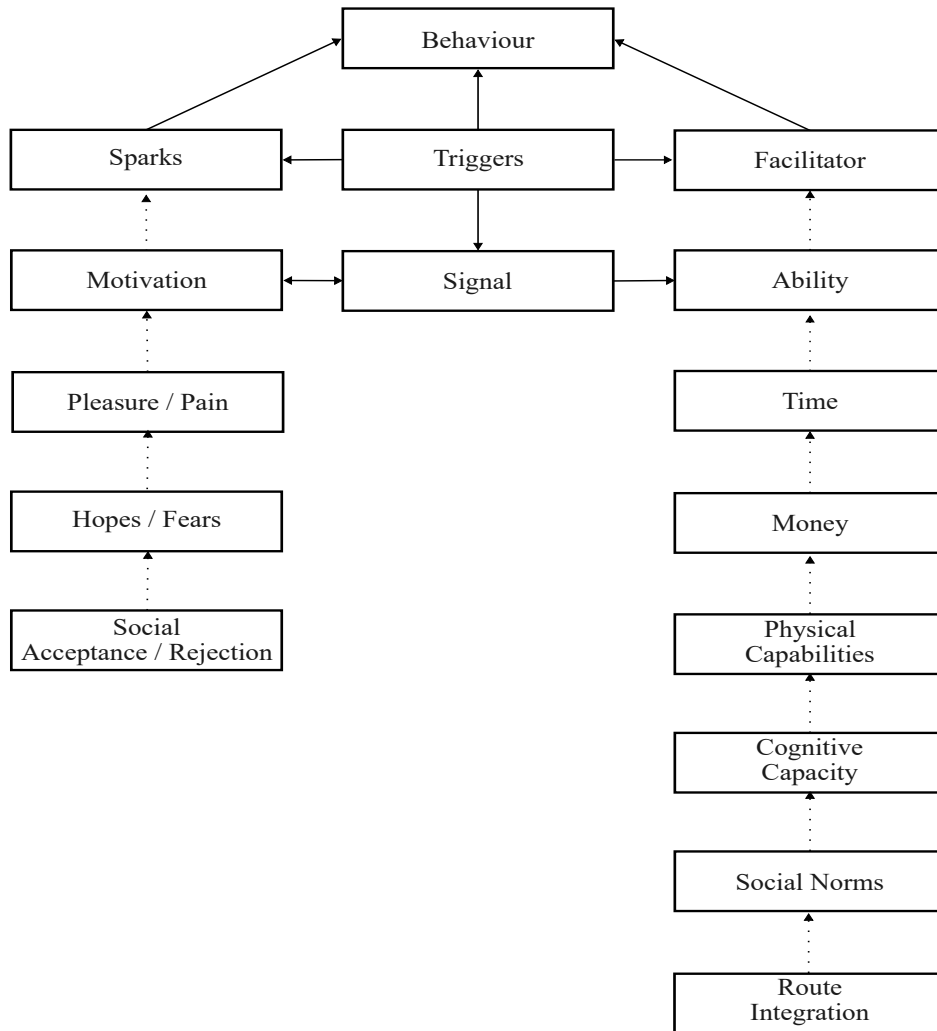


Figure 1: Fogg Behavioural Model (FBM) [10]

Spark triggers are designed to increase low motivation through the aforementioned considerations of emotions and society-based standings. These have largely been implemented through text-based reasoning or video channels that aim to change perceptions about existing behaviour. An example intervention includes positive or negative framing of the consequences for not performing the intended behaviour.

Facilitator triggers are used to increase low ability through emphasizing the simplicity of an activity. Often, the required resources are also highlighted to demonstrate that the user has everything required to succeed in their goals.

Lastly, signal triggers are used as the optimal method of communication when motivation and ability are high. There are no specific design techniques that are leveraged as a simple notification (e.g., reminder) will suffice to drive behaviour.

2.7 Persuasive System Design Model

The Persuasive System Design (PSD) model [11] (Figure 2) is an effective psychology-based theoretical framework aimed at changing a user's attitude or behaviour. It focuses on leveraging 28 identified design rules to implement or evaluate a persuasive system, consisting of three components; an intent, event, and strategy.

Firstly, the intent focuses on understanding the user and system requirements in order to change an attitude towards a desired behaviour. The roles of the system and its end-goals must be assigned. In mapping out the system, inconsistencies in intervention design can also be identified and rectified. Additionally, triggers for behaviour change can be developed to execute at the opportune time.

Secondly, the event focuses on identifying user contexts. This helps map out the key attributes that are associated with the domain of interest. For example, identifying common attributes which lead to negative performance and creating an intervention to promote positive behaviour with actionable, simple implementations. The user context must consider preferences, goals, performance history, and self-efficacy in order to assign reasonable milestones.

Lastly, strategy focuses on improving the communication between the persuader and persuadee. Inconsistencies are identified and removed from the primary message and the intervention channel is selected.

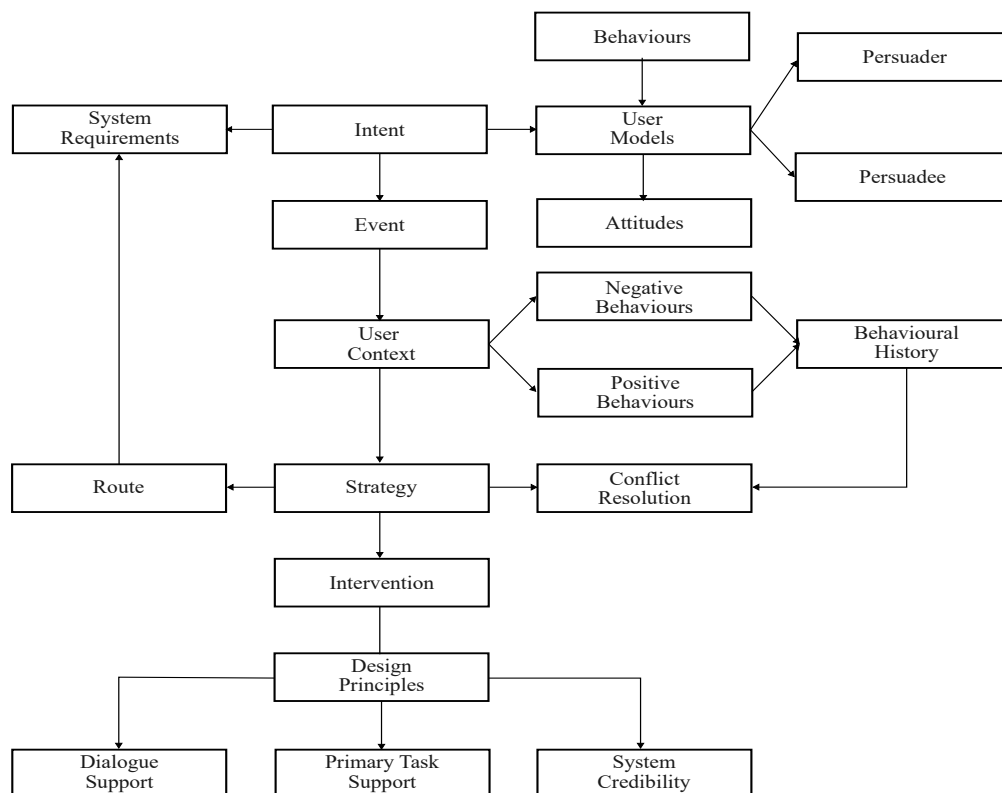


Figure 2: Persuasive System Design (PSD) Model [11]

2.8 Social Cognitive Theory of Self-Regulation

The Social Cognitive Theory (SCT) [12, 13] (Figure 3) of Self-Regulation is a socio-psychological framework by Bandura that explains human behaviour through the processes and consideration of self-monitoring, personal judgment, environment, and self-reflection. It has widely been used in areas such as clinical education and healthcare [13]. It posits that human behaviour is caused by a mixture of internal and external circumstances that require consideration in order to motivate a desired behaviour. As a result, behaviour change requires a multi-level approach which consists of considering personal and environmental circumstances.

Self-monitoring refers to how performance metrics of behaviour affect continued motivation and persistence. In particular, it is heavily affected by psychological factors such as self-esteem and emotional states. Through self-monitoring, an individual can gain insights into what kinds of behavioural patterns drive success and define tangible goals to achieve in the future. Goal setting, if used properly through self-monitoring, can increase self-esteem by highlighting successes and identifying achievable corrective changes in behaviour from failures.

Judgements of successes and failures which result from self-monitoring are defined intrinsically and extrinsically based on personal and social constructs. For example, intrinsic judgments are shaped based on how people have reacted to their behaviour in the past whereas extrinsic judgments depend upon an individual's performance with relation to their peers. Each of these judgments will be defined by a particular environment and the opportunities that it provides, forming outcome expectancies. Through judgments, self-reactive influences can take form which drive behaviour change through perceived urgency. Urgency is often defined through self-satisfactions that serve as a reward function.

Self-reflection, through goal-setting and rewards operate under a negative feedback model where discrepancies between behaviour and goals are reduced through self-introspection. A feedback loop of self-introspection is implemented by aiming to increase ability and self-efficacy through new goal-setting.

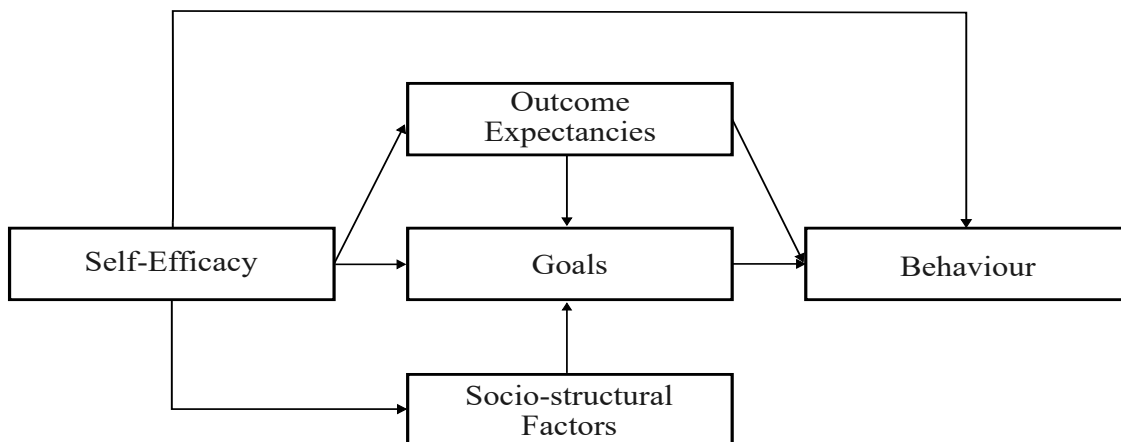


Figure 3: Social Cognitive Theory of Self-Regulation (SCT) [12, 13]

2.9 COM-B Model

The Capability, Opportunity, and Motivation (COM-B) model [14] (Figure 4) is a systematic framework that is used to describe behavioural systems from a psychological perspective. Unlike other models, it incorporates the intervention type, target population, and context which heavily influences system outcomes. Furthermore, unlike most of the existing behavioural models, COM-B is grounded in behaviour change techniques and frameworks [14] that have been validated and studied in clinical trials [15].

Capabilities represent the ability of an individual to achieve a desired behaviour. It is considered at a physical and cognitive level. For example, physical considerations can be related to mobility or pre-existing health-conditions. Cognitive considerations are related to the ability to engage in logical reasoning that motivates behaviour change. This requires prerequisite knowledge about the behavioural domain and required skills to perform a behaviour. The ability to think and reason through the completion of a specific behaviour above all else in order to obtain a positive metric is required to drive motivation.

Opportunities represent the sum of external factors from the individual which affect behavioural outcomes. Such considerations can include factors such as the environment, culture, and social networks of the individual.

Lastly, motivation represents the cognitive considerations from capabilities that lie outside of goal-setting and reasoning. For example, these processes can often be ascribed to emotion-based responses, impulses, operant conditioning, and analytical decision making.

In order for a specific behaviour to occur, each of the three components must be sufficiently high, otherwise, there would be multiple competing behaviours to choose from. Furthermore, each of the components directly influences each other, meaning that a deficiency in one factor may be supported by another one. This is best represented by the double-headed arrows in Figure 4 connecting each of the conditions to behaviour, forming a strongly connected graph.

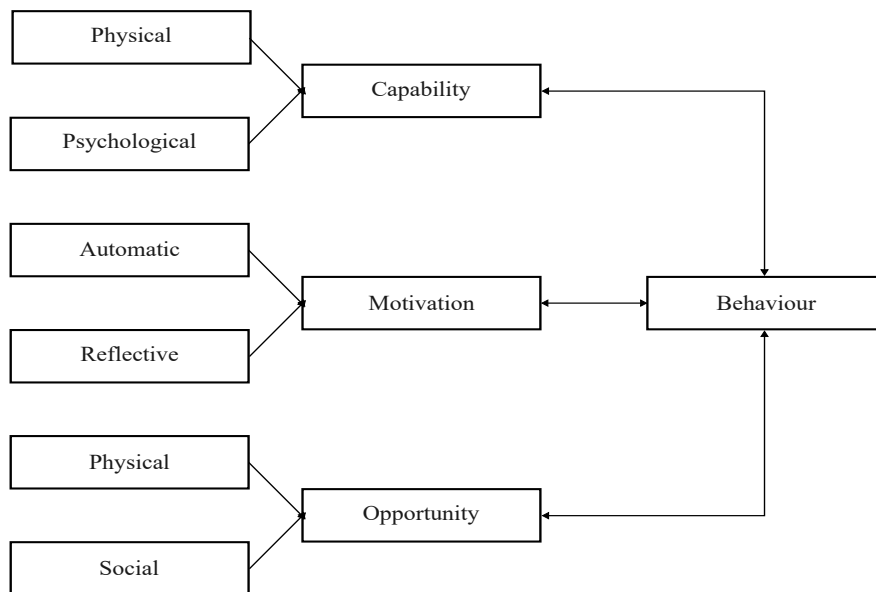


Figure 4: Capabilities, Opportunities, and Motivation (COM-B) [14, 15]

2.10 ComTech Model

ComTech [1] is a unified psycho-technical framework that amalgamates design models grounded in communication design (socio-psychology), user experience design (technology acceptance), and persuasive technology (behaviour change) together. It aims to provide researchers with tools to help change behaviour in domains such as healthcare by providing a robust multidimensional resource.

This thesis leverages an adapted ComTech Model (Figure 5) which only focuses on user experience (technology acceptance) to address the third research question from Chapter 1.3, “How does user experience (e.g., trust, privacy, and perception) affect user attitudes and adherence to fitness-based recommender systems?” The systems are evaluated from a perception and behaviour-based approach.

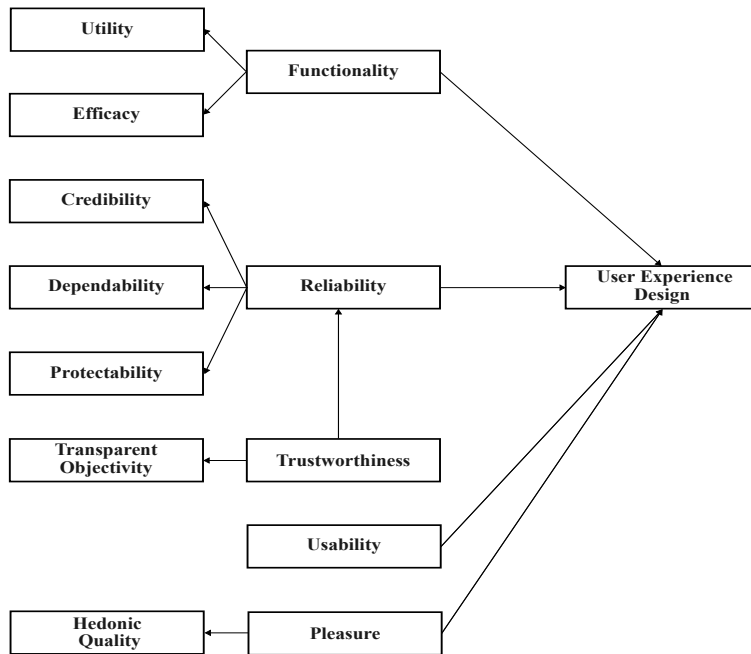


Figure 5: Adapted HUED-CUP Framework [1]

To this end, to adapt the original model, user trust was added as a component to the seven-item list of supported qualities (HUED CUP). Each of the supported qualities are summarized in Table 7.

Table 7: User Experience Design Techniques from the Adapted ComTech Model [1]

Technique	Definition
Utility	The extent to which the intervention matches a user’s requirements and expectations.
Efficacy	The extent to which the intervention is able to complete tasks to an acceptable degree.
Credibility / Dependability	The extent to which the intervention can be relied upon.

Table 7 (continued)

Technique	Definition
Protectability	The protection of user data from abuse, compromise, or inaccessibility.
Usability	The simplicity of an intervention requiring minimal effort and difficulty to understand or use.
Perceived Persuasiveness (Hedonic Quality)	The ability of the intervention to persuade a user based on aesthetics.
Transparent Objectivity	The extent to which the outputs of the intervention are explainable, impartial, free of bias, and objective.

2.11 Theory of Interactive Media Effects Model

The Theory of Interactive Media Effects (HAI-TIME) [16] (Figure 6) model is a psychology-driven framework incorporating privacy and user control into the design and evaluation of AI-driven systems. It leverages a cue and action-based route to maximize user experience and system perception by supporting transparent processes and minimal threat to personal identity. The goal is to support a robust human-machine collaboration through equal participation in decision making and intelligence amplification.

The cue-based route uses user interface heuristics to trigger fast positive perceptions about the systems credibility. For example, negative consequences caused by the machine heuristic (e.g., mechanistic, cold, unyielding) can be minimized through supporting social presence and authority-based heuristics (e.g., Bandwagon [74]) and explanations.

The action-route increases user trust by supporting a transparent human-machine collaboration. In particular, the user may choose from one of many routes to obtain their desired goal that best aligns with their personal preferences, increasing agency and affordance. A common example involves having users explicitly opt in or out of data collection policies [16]. In general, the action route is implemented through user interface elements that aim to anthropomorphize its outputs.

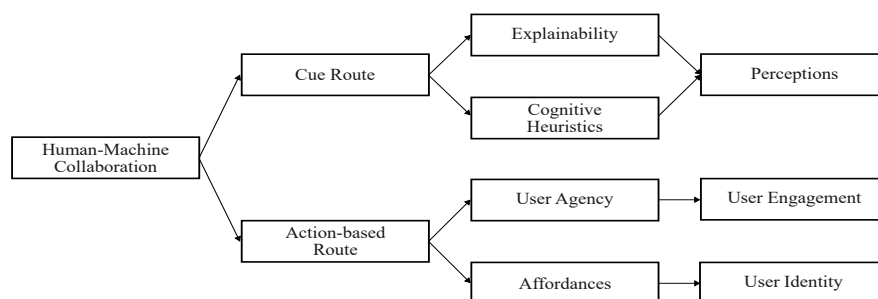


Figure 6: Theory of Interactive Media Effects (HAI-TIME) Model [16]: Cues and Action Routes

3 Methodology

This work used the Design Science Research Methodology (DSRM) Process Model [17], Rapid Prototyping [17, 20], and a mixed methods study design [95, 2] to address the research questions outlined in Chapter 1.3. The dataset, software tools, workflows, participant information, and statistical tests that we used are discussed and reported in the following chapter.

3.1 Design Science Research Methodology

This project used the Design Science Research Methodology (DSRM) Process Model [17] (Figure 7) to conceptualize and evaluate the interventions through a user and computational study. The DSRM Process Model was selected as it balances design science with engineering requirements whereas more traditional frameworks such as the System Design Framework are less adaptable [17]. It also also been used in related persuasive technology research in the past [19]. The conceptualization, implementation, and steps to carry out the two studies strictly followed from the sequential steps of the DSRM Process Model that are outlined in the following chapters. The DSRM Process Model is a specialized six-phase process for carrying out research in information systems. The methodology enables a feed-forward and cyclic workflow that promotes iterative improvements. The dotted lines in Figure 7 represent the backpropagation. The six phases consist of problem discovery, brainstorming, implementation, formative evaluation, and the dissemination of findings.

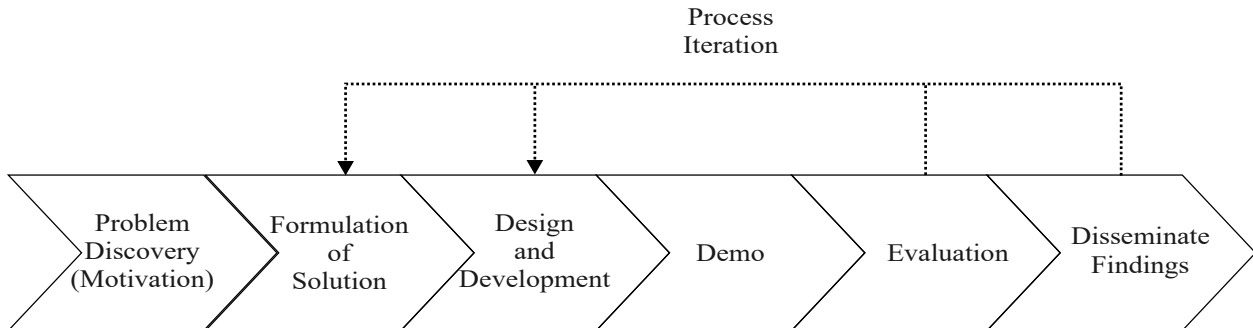


Figure 7: Design Science Research Methodology (DSRM) Process Model [17]: Problem Discovery & Motivation, Formulation of Solution, Design & Development, Demo, Evaluation, Dissemination of Findings

3.1.1 Phase 1: Problem Discovery & Motivation

The first phase of the study consisted of identifying limitations from existing literature and justifying the urgency for this work through a systematic literature review [28]. Chapter 2 identified and summarized the limitations of existing user and computational studies, ultimately justifying the need for personalized, trustworthy, and explainable fitness-based recommender systems.

3.1.2 Phase 2: Formulation of Solution

The second phase consisted of identifying a methodological approach to carry out the design and evaluation of our computational and user study. To this end, we leveraged a unified framework (Figure 8) grounded in psychology and behaviour-driven systems to facilitate the design, implementation, and evaluation of our fitness-based recommender systems and user study outcomes. The structure of the applied framework is heavily grounded in Bandura’s framework for reciprocal determinism [18].

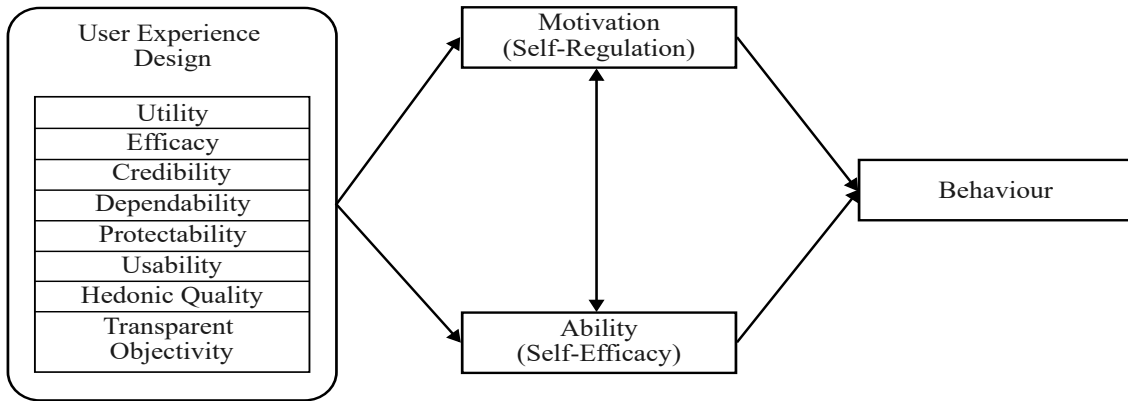


Figure 8: Applied Framework for Projected Behaviour Performance Adapted from [1, 2, 18, 19]. Elements of the Framework are Leveraged from FBM [10], PSD [11], SCT [12], COM-B [14], HUED-CUP [1], and HAIL-TIME [16]

Each of the constructs were evaluated based on Likert scale responses and statistical analyses which is discussed in Chapter 3.1.5. Table 8 outlines how the related entities and constructs of Figure 8 have been implemented into the system design.

Table 8: Explanation of How Entities and Constructs from Figure 8 Have Been Considered in System Design

Technique	Definition
Utility	The features of the content-based and context-aware recommender systems were selected based on common trends and gaps in the literature. Chapter 3.1.1 describes the process in greater detail.
Efficacy	The content-based recommender system intentionally used drop-down spinners to select features which reduces the likelihood of human error during the recommendation generation. The context-aware recommender system leveraged simple features with plain language to make the interaction straightforward.

Table 8 (*continued*)

Technique	Definition
Credibility / Dependability	Credibility and dependability were supported through explainable feature relevance bar plots of the classifier models and having the LLMs act as professional bariatric healthcare experts. Furthermore, the Agents thought process was logged and viewable to the user to determine how credible the final recommendation was.
Protectability	Protectability was implemented through the privacy policy of the recommender systems.
Usability	The user interface elements were carefully chosen and selected to maintain minimalistic design. Widgets and elements are only included if they were necessary to facilitating communication between the system and user.
Perceived Persuasiveness (Hedonic Quality)	Persuasiveness was supported through text-based reasoning, explainability, and the privacy policy. The systems which rely on persuasive elements leveraged psychosocial theories in the prompt through explicit instruction or retrieval augmented generation. This had an impact on how much self-efficacy or self-regulation was prioritized in the final recommendation. The Agent had full autonomy to select the persuasive design frameworks to utilize based on contextual factors such as the environment and capabilities. Hedonic Quality was supported through stream-based chat systems to avoid the potential of a cold machine heuristic [16]
Transparent Objectivity	Transparent objectivity was supported through factors related to credibility, dependability, and protectability. In particular, the trust in the system was also supported through being able to view specific Agent tool calls and information from the classifier models.

The unified framework outlined in Figure 8 consisted of conjoined components from FBM [10], PSD, [11], SCT [12], COM-B [14], ComTech [1], and HAI-TIME [16]. Repeated constructs from each of the models were joined together to create a more holistic framework. For example, motivation was defined as using each of the related components from FBM (e.g., Abilities) [10], SCT (e.g., Self-Efficacy) [18], and COM-B (e.g., Motivation) [14] whereas privacy preferences and user control were merged and defined using

components from ComTech (e.g., Trustworthiness) [1] and HAI-TIME (e.g., Explainability, Affordances) [16]. The synthesis was required as no existing frameworks to date have incorporated metrics such as user experience, behaviour, and motivation into a unified model [1].

In conclusion, user experience, motivation, and ability must be sufficiently high to change behaviour. The bidirectional arrow between motivation and ability demonstrates their interrelated dependence on each other. Self-Regulation and Self-Efficacy must work together for the components to activate a behaviour.

3.1.3 Phase 3: Design & Development

The third phase consisted of identifying adequate datasets, models, vector stores, search tools, and evaluation metrics to design and implement the recommender systems. For example, the pipeline to fit each of the machine learning models was designed prior to the implementation phase in order to pre-process the features through augmentation, imputation, one-hot encoding, scaling, and label encoding. Selection of the features for the recommender system was guided by conducting a systematic literature review [28] to identify common features and areas for future work.

The design and development of the user interface was guided by an iterative three-stage rapid prototyping framework, originally proposed by Keyson and Parsons [20]. The framework, shown in Figure 9 was adapted from Willige’s Three Stage Iterative Design Process [21] to work under limited time and resource constraints.

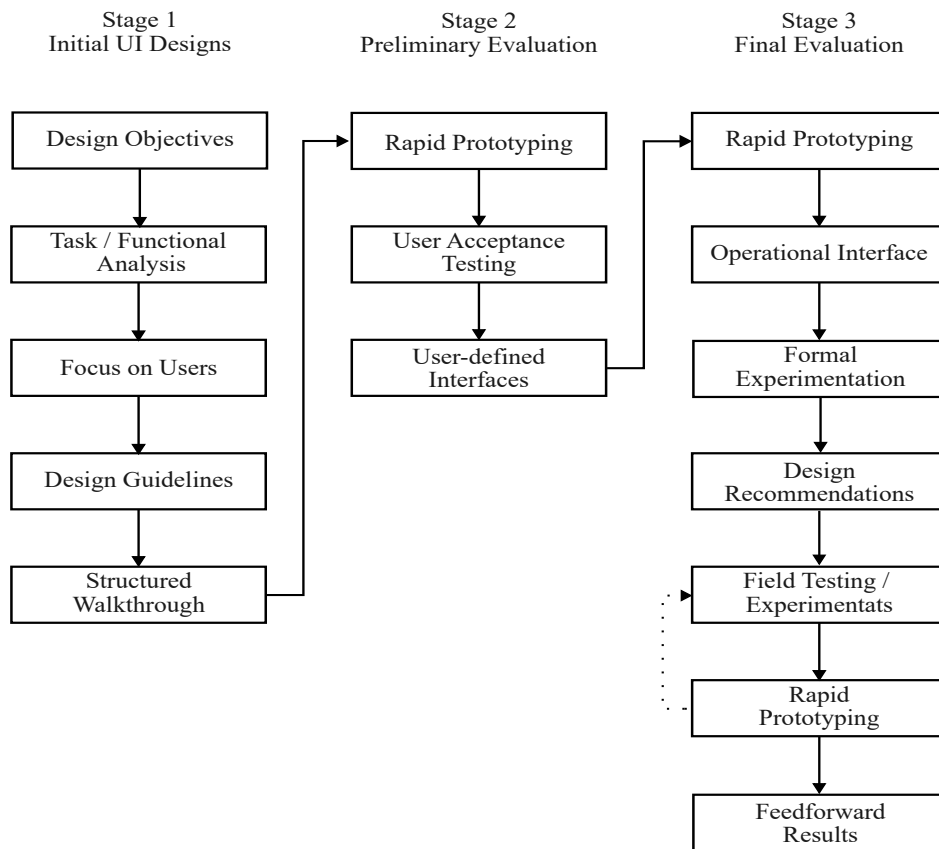


Figure 9: Keyson and Parson’s Adapted Iterative Design Process [20] Adapted from Willege’s Three Stage Iterative Design Process [21]

The first stage consisted of identifying the design requirements of the recommender systems such as the features to make predictions. Additionally, tasks and functional requirements had to be considered to design the manner in which the explainable and privacy-preserving components should be integrated and displayed to the end user. The outputs of the first stage were a preliminary prototype that required user validation.

The second stage consisted of evaluating the preliminary prototype based on informal user acceptance testing and using user interaction. User acceptance testing was informal as ethics clearance was not yet obtained. A list of new changes were aggregated and considered to improve the user interface based on feedback in the third stage.

Lastly, in the third stage, feedback was integrated into the user interface. The operational components were also integrated into the widgets enabling a fully-functioning intervention. Formal experimentation was carried out on users to identify further design recommendations and future needs. Iterative improvements and evaluations were made until the final user interface was designed, implemented, and acceptable for deployment in an experimental user study. Figure 10 shows the finalized user interfaces for the content-based and context-aware recommender system, irrespective of their group.

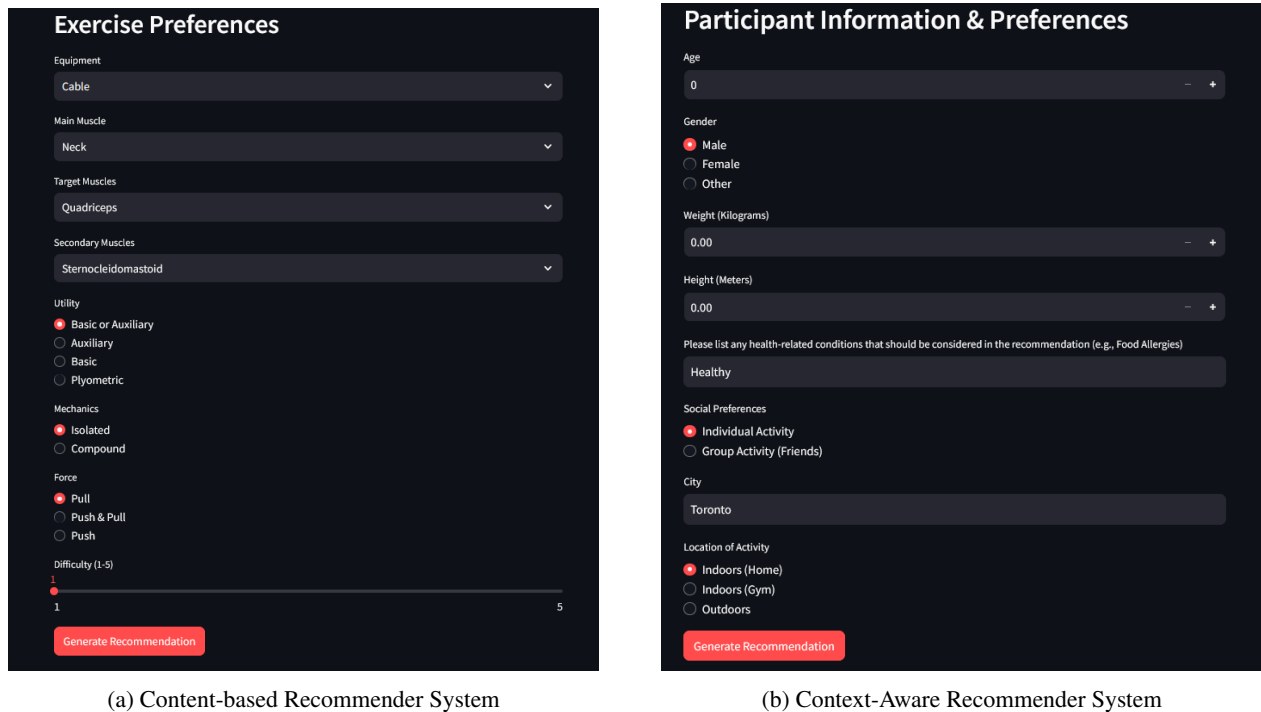


Figure 10: User interfaces for the support recommender systems

3.1.4 Phase 4: Demo

The fourth stage consisted of carrying out a user study that addressed the research questions outlined in Chapter 1.3. The user study aimed to identify how elements of the intervention such as explainability, trust, and threat to identity affected perceived user experience, perception, and adherence to fitness-based recommender systems using qualitative and quantitative metrics.

The experimental group leveraged clear and transparent privacy policies, open-source models, and interpretable visualizations that demonstrated how the user’s information was used to generate the exercise recommendations. Figures 11 and 12 show examples of the explainable metrics supported in the experimental content-based recommender system.

Text-based reasoning was used to explain why the recommendations were beneficial to perform based on theoretical psychosocial frameworks grounded in the literature. Figures 13, 14, 15 and 16 show examples of the persuasive text-based reasoning elements supported in the experimental content and context-aware recommender systems fitness and diet plans, as generated by the large language model.

Based on the provided inputs, we suggest the following exercises with their corresponding relevance:

Recommendation	Relevance
Neck Flexion	45.0%
Neck Extension	9.0%
Lateral Neck Flexion	8.0%
Shoulder Press	5.0%
Chin-up	4.0%

Figure 11: Relevance of Exercise Recommendations from the Experimental Content-Based Recommender System

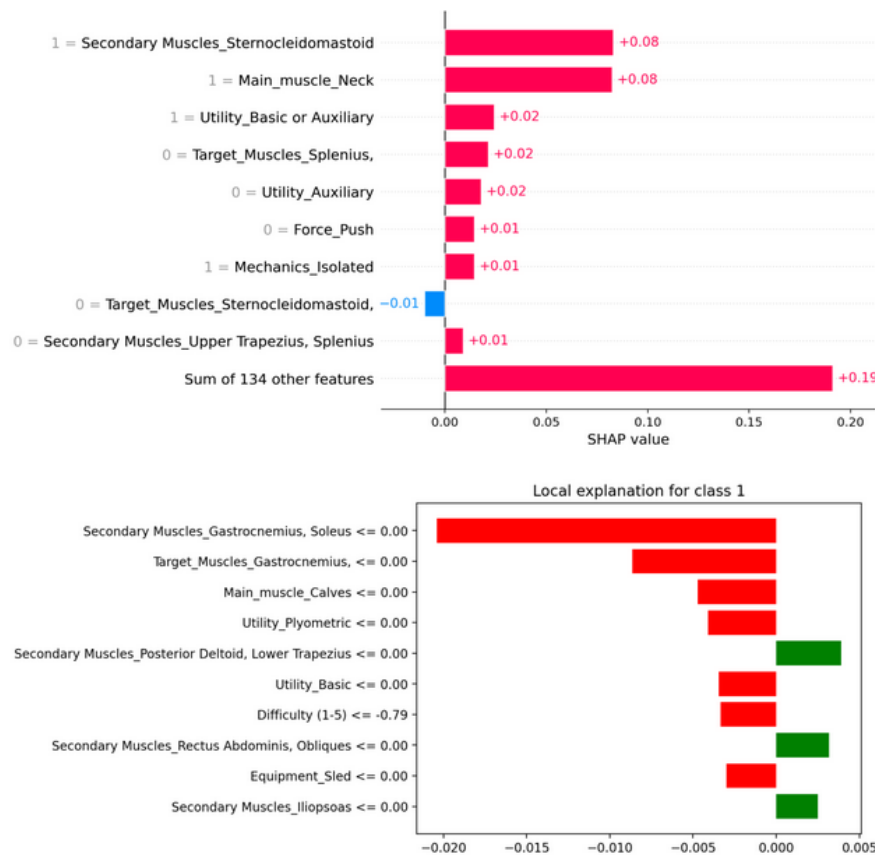


Figure 12: Explainable Feature Relevance Plots from the Experimental Content-Based Recommender System

Exercise Plan

Neck flexion is an incredibly beneficial exercise for maintaining good posture, preventing neck pain, and improving overall neck mobility. Here's why you should incorporate it into your weekly schedule:

Benefits:

1. **Improved Posture:** Neck flexion helps strengthen the muscles in the front of your neck, which can help counteract the effects of prolonged sitting or looking down at a phone or computer. By strengthening these muscles, you can help maintain a neutral neck position and prevent the 'forward head posture' that can lead to poor posture and neck pain.
2. **Neck Pain Prevention:** Regular neck flexion exercises can help prevent and alleviate neck pain. By strengthening the muscles in your neck, you're providing better support for your head and reducing the strain on your neck and upper back.
3. **Improved Mobility:** Neck flexion helps increase the range of motion in your neck, allowing you to turn your head more easily and reducing stiffness. This can be particularly beneficial for those who spend a lot of time driving or working at a desk.

Figure 13: Exercise Plan Generated by the Experimental Content-Based Recommender System

Incorporating into Weekly Schedule:

- **Motivation:** Set a reminder on your phone or computer to do neck flexions every hour. You can also make it a habit to do them whenever you take a break from work or check your phone.
- **Ability:** Start with small, manageable sets. Aim for 10-15 repetitions, 2-3 times a day. As you get stronger, you can increase the number of sets or repetitions.
- **Triggers:** Pair the exercise with other daily habits. For example, you could do neck flexions every time you make a cup of tea or coffee, or every time you sit down at your desk.

Here's a simple way to perform a neck flexion:

- Sit or stand with your back straight.
- Slowly lower your chin towards your chest, keeping your shoulders relaxed.
- Hold for a few seconds, then slowly return to the starting position.
- Repeat for the desired number of repetitions.

Figure 14: Persuasive Techniques to Increase Exercise Motivation in the Experimental Content-Based Recommender System

"Great question! Given your BMI of 35.16, you're in the obese category, and it's crucial to approach weight loss in a healthy and sustainable way. Here's a simple, yet effective plan using the Fogg Behavior Model to help you get started:

1. **Motivation:** Remember, every small step counts. Celebrate your progress, not just the end result. You're doing this for your health and well-being, and that's something to be proud of!
2. **Ability:** Let's make this easy and enjoyable. Since you prefer individual activities indoors, consider a gym membership. Here are some exercises tailored for your condition:
 - o **Cardio:** Start with low-impact exercises like swimming or cycling. Aim for 150 minutes of moderate-intensity or 75 minutes of vigorous-intensity activity per week.
 - o **Strength Training:** Incorporate bodyweight exercises or light weights. Focus on compound movements like squats, lunges, and push-ups. Aim for 2-3 sessions per week.
3. **Trigger:** Set reminders on your phone or use fitness apps to keep you on track. Make exercise a habit by scheduling it like you would any other appointment.

Figure 15: Exercise Plan Generated by the Experimental Context-Aware Recommender System

Now, let's talk about diet. You'll want to focus on whole foods and healthy proteins. Here's a simple guide:

- **Increase:** Fruits, vegetables, whole grains, lean proteins (fish, poultry, beans, nuts), and healthy fats (avocados, olive oil).
- **Decrease:** Processed foods, red meat, cheese, and sugary drinks. Limit your intake of refined carbohydrates and unhealthy fats.

Remember, portion control is key. Aim for 1,500-1,800 calories per day, focusing on nutrient-dense, low-energy-dense foods. Don't forget to stay hydrated!

Lastly, consider keeping a food diary to track your progress and make adjustments as needed. You're not alone in this journey, and every small victory is worth celebrating. Keep going, and you'll see progress in no time!"

Figure 16: Diet Plan Generated by the Experimental Context-Aware Recommender System

The control group used the same classical machine learning models to make fitness-based recommendations, however, explainability was not implemented or considered. Figure 17 shows the contrasting output for the control conditions content-based recommender system.

Based on the provided inputs, we suggest the following exercises with their corresponding relevance:

Recommendation	Relevance
Pull-up	[Hidden]
Curl	[Hidden]
Bench Press	[Hidden]
Shoulder Press	[Hidden]
Triceps Dip	[Hidden]

Figure 17: Relevance of Exercise Recommendations from the Control Content-Based Recommender System

Figures 18 and 19 show examples of the fitness and diet plan recommendations supported by the control content and context-aware recommender systems which leverage no persuasive text-based reasoning elements.

As a friendly bariatric healthcare professional, I'm glad to help you with personalized advice on weight loss through exercise and dietary recommendations. Here's a plan tailored to your profile:

Exercise Recommendations

- 1. Indoor Cardio Workouts:**
 - **Jump Rope:** This is an effective cardiovascular exercise that can be done indoors. Start with 5-minute intervals and gradually increase as your stamina improves.
 - **Stationary Cycling:** If you have a stationary bike, cycling is a low-impact way to burn calories and improve cardiovascular health. Aim for at least 30 minutes a day.
- 2. Strength Training:**
 - **Bodyweight Exercises:** Incorporate exercises like squats, push-ups, and lunges that can be done at home without equipment. These help in building muscle, which can increase your metabolic rate.
 - **Resistance Bands:** If available, use resistance bands for exercises like bicep curls and shoulder presses. These are effective for muscle strengthening and are easy to store.
- 3. Flexibility and Balance:**
 - **Yoga or Pilates:** These can be done indoors and improve flexibility, balance, and core strength. There are many free online resources and videos to guide you through sessions.

Figure 18: Exercise Plan Generated by the Control Context-Aware Recommender System

Dietary Recommendations

- 1. Balanced Meals:**
 - **Portion Control:** Use smaller plates to help control portion sizes and prevent overeating.
 - **Macronutrient Balance:** Aim for a balanced intake of proteins, carbohydrates, and healthy fats. For example, a meal could include grilled chicken (protein), quinoa (carbs), and avocado (healthy fat).
- 2. Increase Fiber Intake:**
 - **Vegetables and Fruits:** Incorporate a variety of colorful vegetables and fruits into every meal. They are low in calories but high in fiber, which can help you feel full longer.
 - **Whole Grains:** Choose whole grains like brown rice, oats, and whole wheat bread over refined grains.
- 3. Hydration:**
 - **Water:** Aim to drink at least 8 glasses of water a day. Staying hydrated can help manage hunger and improve metabolism.
- 4. Mindful Eating:**
 - **Avoid Distractions:** Eat without distractions like TV or smartphones to better recognize your body's hunger and fullness cues.
 - **Chew Slowly:** Take time to chew your food thoroughly, which can aid digestion and help you enjoy your meals more.

Figure 19: Diet Plan Generated by the Control Context-Aware Recommender System

The privacy policy in the control group was clear, stating that closed-source models were used in some of the recommender systems. As a result, their privacy was impacted by third parties privacy policies. Lastly, there was no text-based reasoning to explain the beneficial nature of the recommended exercises.

To summarize, Table 9 outlines the comparisons and differences between each of the functionalities that the four recommender systems support.

Table 9: Comparative Table of the Recommender Systems and Supported Features

Group	System	Privacy	Explainability	Persuasive	Exercises	Diet Plan
Control	Content-Based				✓	
Control	Context-Aware				✓	✓
Experimental	Content-Based	✓	✓	✓	✓	
Experimental	Context-Aware	✓	✓	✓	✓	✓

3.1.5 Phase 5: Evaluation

The fifth stage consisted of carrying out and interpreting the results from our computational and user study to determine how effective the intervention was in reducing sedentary behaviour through robust and persuasive fitness recommendations. The computational study evaluated the effectiveness of recommender systems based on the weighted averages of precision, recall, and F1-score. Furthermore, considerations of computational complexity (e.g., fitting and scoring time) were considered.

The user study employed a 2×2 between-subjects study design resulting in four groups for assigning participants into. The first independent variable assigned the group of a participant; control or experimental. The control group provided recommendations in isolation whereas the experimental group also incorporated explainability, text-based reasoning, and psychosocial frameworks. The second independent variable assigned the type of recommender system that the participant was exposed to; content-based, or context-aware. The effectiveness of the interventions were based upon three dependent variables; perceived user experience, motivation, and projected behaviour performance.

User experience was measured using seven levels derived from the adapted ComTech model [1]. The levels under consideration included perceived trust (transparent objectivity), expertise, privacy, efficacy, utility, usability, and persuasiveness (hedonic quality). Perceived motivation was measured by considering two behavioural determinants; self-efficacy and self-regulation. Lastly, projected behaviour performance was measured by considering self-reported weekly activity and post-intervention weekly activity goals in minutes.

This research design was chosen in order to evaluate qualitative and quantitative aspects of user experience whereas traditional recommender systems research have relied on mostly quantitative, computation-based methodologies and outcome metrics [22, 28]. Furthermore, the study’s research objectives satisfied criteria outlined by Greene et al. [2] for carrying out mixed methods research (See Table 10).

Table 10: Greene et al.'s Criteria for Conducting Mixed-Methods Research [2]

Criteria	Requirement
Triangulation	Requires the results of the study to converge, corroborate, and correspond to two or more different methodologies.
Complementarity	Requires the augmentation of results to elaborate, enhance, illustrate, and clarify study outcomes with relation to two or more methodologies.
Development	Requires the use of two or more methodologies to help inform a study's development such as sampling methods, implementation, and qualitative or quantitative measures.
Initiation	Requires new perspectives and the reformulation of research questions from one methodology to another.
Expansion	Requires the expansion of knowledge from a domain by using different methodologies to address the research questions.

Quantitative factors were obtained through a questionnaire that used Likert scale measures from, “Strongly Disagree” to, “Strongly Agree” and short open-ended responses with numerical validation to record participant demographics. Qualitative factors were obtained from short or long open-ended responses to measure thoughts and perceptions about the usefulness of the recommender systems.

The study was conducted in an in-person controlled laboratory setting to reduce the likelihood of confusion resulting from the questionnaire and intervention instructions. Furthermore, it was necessary to run the study in-person to reduce the likelihood of security issues resulting from the use of API keys.

To this end, the hypotheses with respect to the research objectives were formulated as follows:

1. Explainability of Recommendations

- H_1 : Explainability of the system and fitness recommendations will positively affect user experience outcomes.

2. Psychosocial Theory-Driven Frameworks

- H_2 : The integration of theory-driven social science and psychology frameworks used to design the recommender systems will positively affect projected behaviour performance to reduce sedentary lifestyles.
- H_3 : The integration of theory-driven social science and psychology frameworks used to design the recommender systems will positively affect perceived motivation to reduce sedentary lifestyles.

3. User Experience

- H_4 : User experience will positively affect projected behaviour performance to reduce sedentary behaviour.

Recruitment of participants commenced on April 30, 2025 and continued until June 16, 2025 after obtaining ethics clearance from York University’s Institutional Review Board. Snowball [96] and self-selection [97] sampling were used as both techniques offered cost-effective recruitment strategies for finding a large sample of participants within a short period of time. Table 11 outlines the effectiveness of each recruitment strategy resulting from the participants who attended the in-person study.

Table 11: Applied Recruitment Strategies: Snowballing and Self-Selection Sampling

Recruitment Strategy	Count
Poster	40
E-mail	14
Word of Mouth	10
Reddit	4
Discord	4
LinkedIn	1

One-hundred and twenty-seven people filled in the pre-screening questionnaire to assess for study eligibility. Participants were required to be between the ages of 18-30 years old, having no existing and/or prior physical or cognitive disabilities. 54 participants were excluded from further participation due to failing to meet the eligibility criteria or not responding to the follow-up study invite.

73 participants followed up and attended the in-person appointment. Completion of the study resulted in a monetary compensation of \$20.00 (CAD).

Prior to conducting the user study, ethics clearance was obtained from York University’s Human Participants Review Committee (HPRC). Interested participants were required to fill in a pre-screening questionnaire which assessed for the presence of physical and cognitive impairments. Participants who met the eligibility criteria of being aged 18-30 years old and having no existing physical and cognitive disabilities were invited to take part in the user study.

As participants arrived at the laboratory, they were briefed on the study objectives and prompted to carefully read the informed consent. The experiment began with the pre-intervention questionnaire followed by usability testing and a post-intervention questionnaire. Demonstrations of how to use the recommender system was not provided and assignments into a group and recommender system were handled by two random number generators.

The recommender systems were compiled and run on a Razer Blade (Base 2020) laptop for the participants to interact with. Table 12 summarizes the hardware and software specifications of the live environment.

Table 12: Razer Blade 2020 (Base): Hardware Specification of Laptop Used During Demo

Component	Specification
Operating System	Windows 11 Version 24H2 (x64)
Python Version	3.12.4
CPU	Intel(R) Core(TM) i7-10750H
GPU	NVIDIA GeForce RTX 3060
Storage	2 TB
System RAM	16 GB

Each session of the user study took between 30-45 minutes to complete. Participants were allowed to ask the researcher for help while generating fitness recommendations if the input fields or exercise recommendations were unclear. At the end of the study, participants were remunerated with \$20.00 (CAD) in physical cash.

Table 13 outlines each of the questionnaires and frameworks that were used to design the pre-screening and study questionnaire. Firstly, question ordering among each section of the questionnaire was randomized and half of the original statements were inverted to ensure that the participants were reading and answering the prompts carefully. Secondly, the number of questions adapted from the frameworks was minimized in order to reduce the likelihood of participant fatigue.

The Get Active Questionnaire [4] (Table A.2.1) was used to prescreen participants for risks related to exercise safety and the presence of physical disabilities. Answering, “Yes” to any of the questions indicated a moderate to high risk in performing exercises without the consultation of a health care provider.

The Cognitive Change Questionnaire (CCQ8) (Table A.2.2) [5] was used to prescreen participants for mild cognitive impairments. Answering, “Yes” to any of the questions increased the internal scoring metric by 1. A final score greater than or equal to 2 indicated the possibility of a cognitive impairment. A score greater than or equal to 4 indicated the presence of Dementia.

The Physical Activity Enjoyment Scale (PACES-8) (Table A.3.1) questionnaire was used to assess preliminary fitness motivation among participants prior to interacting with the intervention. It is a shortened version of the original 18-item PACES questionnaire that was developed and comprehensively validated by Kendzierski and DeCarlo [98] to measure psychometric constructs.

Similarly, the Privacy Attitudes Questionnaire (PAQ) [7] (Table A.3.2) was used to measure baseline privacy perceptions and attitudes of personalized technological interventions. Of the 36 original questions, only 9 were adapted to the study questionnaire as they had direct relevance to information and recommender systems.

After participants interacted with one of four possible recommender systems, the study questionnaire leveraged prompts that were motivated by Oyibo’s ComTech Taxonomy [1] (Table A.3.3). In particular, the HUED CUP framework was used to develop 15 questions that assessed user experience. Factors such as perceived credibility, expertise, ease of use, benefit, and trustworthiness were used to develop the questions. The initial framework was adapted to meet the research objectives of this work, re-defining perceived trustworthiness as requiring transparent objectivity and having no bias in the final fitness plan.

Self-efficacy was measured through the Self-Efficacy for Exercise Scale [8] (Table A.3.4). The 9-item questionnaire which adapted its questions from McAuley’s et al.’s work [99] was integrated into the study tool to assess the impact of barriers related to physical activity, self-efficacy, and sedentary behaviour. Following this, open-ended questions were included to determine why the participant made their choices relating to self-efficacy prompts and exercise goal setting.

Self-regulation was evaluated through two questionnaires. Five questions were adapted and used in the study tool from the Exercise Goal-Setting Scale (EGS) [9] (Table A.3.5) and the Exercise Planning and Scheduling Scale (EPS) [9] (Table A.3.5). The questions were selected in order to measure self-regulation with respect to multiple factors such as social support, self-efficacy, and outcome expectations. Similarly, an open-ended question was included to determine the rationale behind the participant’s selections.

Lastly, demographics of the participants were captured at the end of the study. Gender, age, ethnicity, weight, height, university major, and self-reported physical activity (in minutes) per week were recorded. It was necessary to capture information relating to biomarkers and fitness levels in order to determine whether the intervention had more or less success and benefit with respect to different groups of participants. The consent form (Section A.1), pre-screening questionnaire (Section A.2), and study questionnaire (Section A.3) are available in the appendix.

Table 13: Adapted Questionnaires Used to Evaluate the Fitness-Based Recommender Systems

Source	Assessment	Scale	Reference
Get Active Questionnaire	Physical Activity Safety	Yes / No	Petrella et al. [4]
Cognitive Change Questionnaire	Cognitive Impairment	Yes / No	Damin et al. [5]
ComTech Taxonomy	User Experience (UX)	1 - 5	Oyibo [1]
Exercise Goal-Setting Scale (EGS)	Self-Regulation	1 - 5	Rovniak et al. [9]
Exercise Planning & Scheduling Scale (EPS)	Self-Regulation	1 - 5	Rovniak et al. [9]
Privacy Attitudes Questionnaire (PAQ)	Privacy	1 - 5	Chignell et al. [7]

Table 13 (*continued*)

Source	Assessment	Scale	Reference
Physical Activity Enjoyment Scale (PACES-8)	Exercise Motivation	1 - 7	Mullen et al. [6]
Self-Efficacy for Exercise Scale (SEE)	Self-Efficacy	1 - 7	Resnick and Jenkins [8]

Table 14 outlines the statistical analyses that were undertaken on participant data with respect to the resources used to obtain the results. Quantitative analysis was used to measure reliability tests on the measured constructs and to identify statistically significant differences and correlations between intervention groups with respect to user experience, self-regulation, privacy attitudes, exercise motivation, and self-efficacy.

Qualitative analysis was used to identify significant differences between intervention groups with respect to sentiment analysis of open-ended questions relating to self-efficacy, self-regulation, and user experience. Sentiment classification was compared against the author’s coding and twitter-roberta-base-sentiment-latest [100]. The Transformer model was used due to limited resources and time constraints.

Table 14: Software Used for Statistical Analysis

Software (Version)	Analysis	Assessments
Jasp	Quantitative	Descriptive Statistics Chi-Square Test McDonald’s Omega Spearman’s Rho Post-Hoc Power Analysis
R (Ver 4.4.2)	Quantitative	Aligned Rank Transform ANOVA Post-hoc Contrast Testing (ART-C)
Python (Version 3.13.3)	Quantitative	Cohen’s Kappa
twitter-roberta-base-sentiment-latest [100]	Qualitative	Sentiment Analysis Classification

3.1.6 Phase 6: Dissemination of Findings

The last stage consisted of effectively communicating and disseminating the obtained knowledge from this thesis study. To this end, the results, key limitations, future work, and proposed interventions have been made available through journal and conference publications (e.g., Toyonaga and Oyibo [28]), YouTube videos, York University Dataverse, and York University Libraries (eTheses and Dissertations).

3.2 Recommender Systems Development

This chapter outlines how each of the recommender systems were developed to support the generation of personalized and tailored fitness recommendations.

3.2.1 Hardware

The computational experiments were conducted using Google Colaboratory; a cloud-hosted Jupyter Notebook environment. Table 15 summarizes the hardware and software specifications of the free-tiered environment.

Table 15: Google Colaboratory: Hardware Specification

Component	Specification
Enterprise Version	Free
Operating System	Ubuntu 18.0.4 LTS
Python Version	3.10.12
CPU	Intel(R) Xeon(R) CPU (2.20 GHz)
GPU	Tesla T4 (15GB VRam)
Storage	107.7 GB
System RAM	16 GB

3.2.2 Content-based Recommender System

Murarka’s gym exercise dataset [3] was used to train and evaluate the content-based recommender system. It consists of 340 unique exercises, 617 rows of data, and 17 columns that comprehensively outline the target muscle groups, required equipment, completion difficulty, and text-based descriptions of how to perform the exercise. The dataset was curated by webscraping ExRx.net, a website tailored for fitness professionals and coaches.

Eight columns were dropped from the dataset as they were either irrelevant to content-based features or consisted of mostly null values. Table 16 outlines the features and label that were used to train the machine learning model which supported the recommender system. Given the relevant features, the exercise name column was used as the label for multi-class classification.

Table 16: Murarka’s Gym Exercise Dataset [3]

Attribute	Description
Exercise Name	The recommended exercise (label), resulting from the features that were passed to the model.
Equipment	The required equipment, accessories, or resources required to perform the exercise.
Utility	The importance of the exercise within the context of a fitness plan. Utility can be basic, auxiliary, or plyometric.
Mechanics	The type of joint movements required to complete the exercise. Mechanics can be isolated, only requiring the movement of one joint at a time, or compounded.
Force	The direction of force that is exerted on a target muscle group. Force directions can require that the muscle be pushed, pulled, or both.
Main Muscle	Describes the main muscle group that the exercise targets. It is a higher-level categorical variable in comparison to the target or secondary muscles. Examples include the forearm, upper arms, or chest.
Target Muscles	The primary muscle that is targeted and stressed during the exercise.
Secondary Muscles	The group of muscles that are exerted in addition to the target muscle group during an exercise.
Difficulty	A Likert scale rating assigning the labour and complexity required to complete the exercise. Values are interpreted on an ordinal scale from 1 representing beginner-friendly to 5 denoting extremely challenging.

Seven machine learning models of varying complexity were selected as candidates to support the content-based recommender system: Logistic Regression, Perceptron, Decision Tree, Random Forest, Gradient Boosting Trees, Multilayer Perceptron, and XGBoost. Each model was trained and evaluated using weighted precision, recall, and F1-score from sci-kit learn’s classification report to determine the best model. Cross-

Validation could not be performed due to the small size of the dataset.

The complete training and evaluation pipeline is visualized in Figure 20. The training, testing, and validation sets were generated with a 60:20:20 ratio. Two pipelines for numerical and categorical features were established to manage data pre-processing. In the numerical pipeline, missing values were replaced using Univariate imputation [101] with a mean strategy. The data was subsequently scaled using z-score normalization [101]. Following this, in the categorical pipeline, missing values were imputed using Univariate imputation by considering the most-frequent item approach. One-hot encoding [101] was applied, ignoring unknown categories.

Due to high class imbalance in the dataset, labels with a count less than 5 were removed to reduce noise among the fitting and evaluation process. Additionally, data augmentation was applied using non-majority Random Oversampling [102] on the training set as there was an insufficient number of data points. In each split, the labels were stratified to reduce the likelihood of sampling errors and high variance. Figure B.1 shows the class balance after cleaning the data.

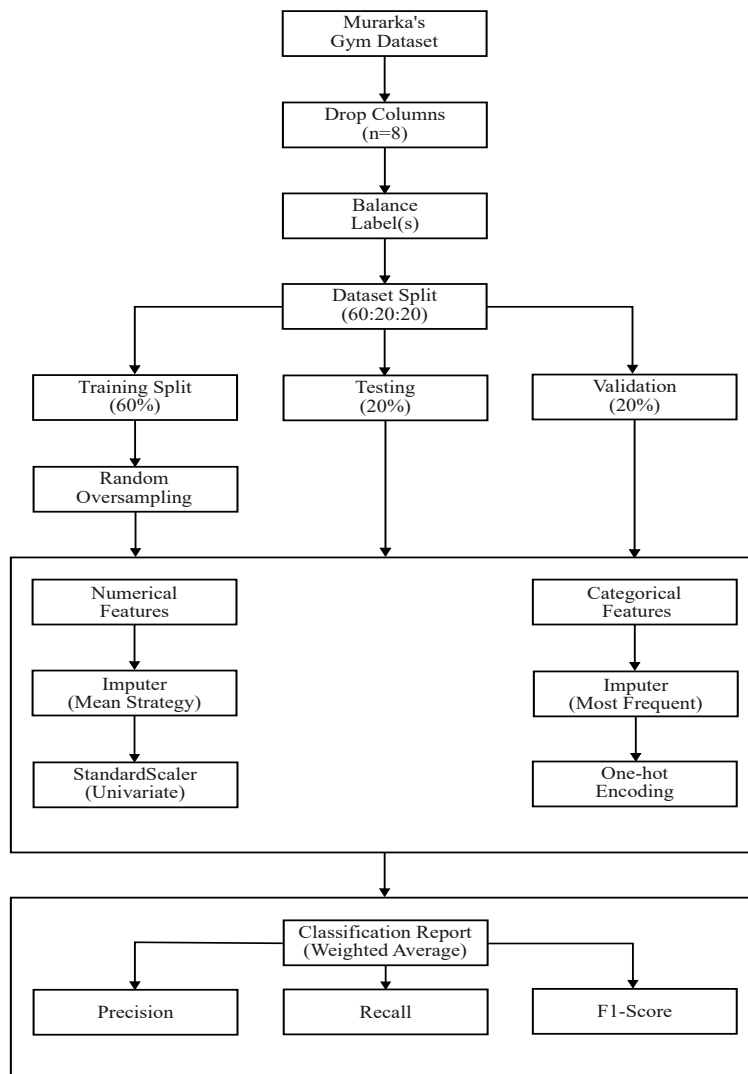


Figure 20: Content-based Filtering Recommender System: Training and Evaluation Pipeline

3.2.3 Context-Aware Recommender System

A LangChain Zero Shot React Description Agent [103] was used to support the context-aware recommender system. It leveraged state-of-the-art search tools and retrieval augmented generation (RAG) to curate personalized fitness plans. Given the user’s demographic and contextual features, the information was formatted into a task-oriented prompt (See Figure C.1). The difference in prompts between the control and experimental group are in how the “question” variable was integrated to support persuasive recommendations (See Figure C.2 for the control question and Figure C.3 for the experimental question). Through the prompt, the agent was able to reason about and generate a comprehensive fitness plan by considering possible actions-state pairs and observations resulting from tool calling. The verbose setting of the agent was enabled to follow its thought process and add explainability to the system.

Table 17 outlines the features that were requested to generate the prompts.

Table 17: Context-Aware Recommender System: Required Features

Attribute	Description
Age	The age of the user (in years).
Gender	The gender of the user (Male, Female, Other).
Weight	The weight of the user (in kilograms).
Height	The height of the user (in meters).
Health Conditions	A list of health conditions or dietary preferences to incorporate into the fitness plan.
Social Preferences	Whether the user prefers to perform exercises in a group activity or by themselves.
City	The city where the user is located.
Location of Activity	Whether the user would like to perform their exercise at home, a gym, or outdoors.

The agentic workflow that supports the control and experimental group of the context-aware recommender system is shown in Figure 21. The dotted lines represent functionalities that are only supported in the experimental group.

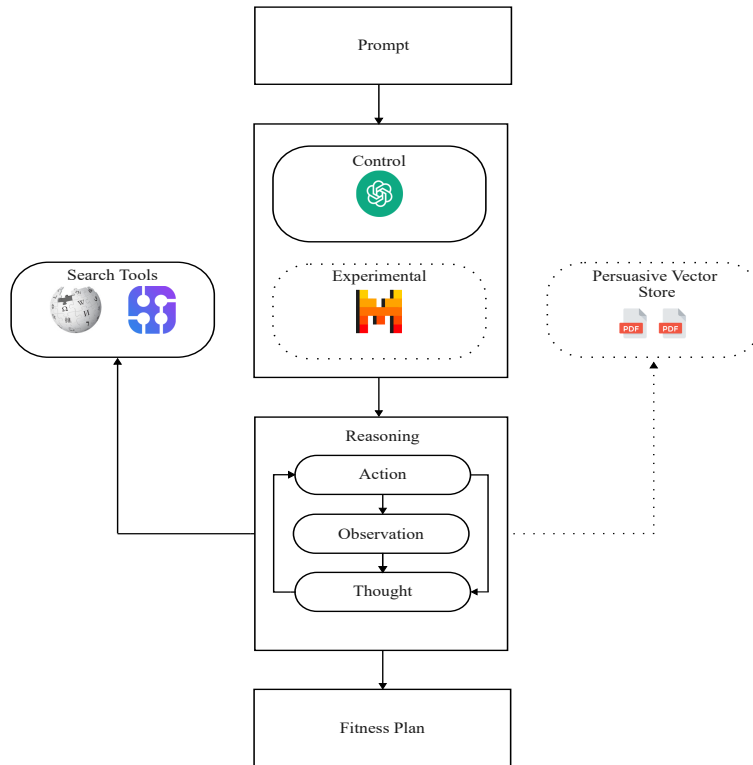


Figure 21: Context-Aware Recommender System: Agent Workflow

The control group leveraged a closed-source OpenAI large language model whereas the experimental group used an open-source instruction-tuned large language model with significantly less parameters. The comparisons of each agent’s configuration is summarized in Table 18.

The vector store that was included in the experimental group only contained two documents (PDFs) by Fogg [10] and Oinas-Kukkonen and Harjuma [11]. As a result, the cues which added persuasiveness to the recommendations were only supported by FBM and PSD. The default sentence-transformers/all-mpnet-base-v2 model was used to embed the documents into a vector space for semantic similarity search during run-time.

Both groups used LangChain’s self-correcting parsing capabilities in the event that an incorrect response structure was generated by the large language model. This also prevented the agents from getting stuck in a permanent thought loop where no existing action-state was properly called or considered. The selections for model temperature was based on repeated trial-and-error until the model produced satisfactory recommendations.

Table 18: Agent Configuration

Agent	Tools	LLM Settings
mistralai/Mistral-Nemo-Instruct-2407 (Zero Shot React Description)	SERP API() Wikipedia VectoreStoreIndexCreator() HuggingFaceEmbeddings()	Temperature (0.6) max_new_tokens (4000)
gpt-4o (OpenAI_Multi_Functions)	SERP API() Wikipedia	Temperature (0.5) max_retries (5)

3.3 Explainability & User Experience

Explainability was implemented in the experimental group using two modalities; text-based reasoning and data visualization. In the experimental group, the content-based recommender system leveraged a Random Forest Classifier which was used to predict suitable fitness plans.

Firstly, text-based reasoning was supported through the presentation of the 5 top classification predictions using sci-kit learn’s predict_proba method. The method returned the top most likely labels with their probabilities, given a set of features. The open-source large language model, mistralai/Mistral-Nemo-Instruct-2407 was used to generate a persuasive rationale for adhering to the fitness plan by leveraging FBM [10]. FBM played a key role in having the large language model integrate the exercises into a weekly schedule. The prompt to support the text-based reasoning is provided in Figure C.4.

Secondly, Chain-of-Thought prompting (CoT) was used in the experimental context-aware recommender system to identify how the agent identified a relevant fitness plan. The verbose setting was enabled to allow the user and developer to see how the agent thinks and calls tools at each step of the reasoning process.

Lastly, data visualization methods were only applied in the experimental content-based recommender system. SHapley Additive exPlanations (SHAP) was used with a TreeExplainer to generate a bar plot that visualized global feature relevance for each prediction. Similarly, Local Interpretable Model-Agnostic Explanations (LIME) was used with a TabularExplainer to provide local feature relevance with respect to the predicted labels.

In supporting text-based reasoning and explainability in the experimental group, users could see why certain features were requested to make an exercise recommendation in addition to their contribution to the predicted labels.

3.4 System Design

In this section, the high-level system design is briefly summarized. The user interface was implemented using the open-source Python library, Streamlit. A multi-page structure was chosen to organize each of the recommender systems into a vertically-aligned navigation bar. Each group had an introductory page that defined open or closed-source terminology and the types of recommender systems followed by a general

privacy statement (See Figure D.1 for a screenshot of the control group’s privacy policy and Figure D.2 for the experimental group). Participants were then guided to the proper intervention to generate recommendations.

A spinner was used when generating recommendations with the classical machine learning model (Random Forest) and agent to communicate that the task was being processed in real-time. The time to produce a comprehensive recommendation or explanation ranged from 20 to 60 seconds due to the complexity of the prompts and search strategies of the agent. Lastly, the final response was incrementally streamed to the user in markdown format to minimize the chance of inadvertently creating a negative machine heuristic [16].

Model parameters were all set to verbose mode in addition to logging messages at each step of the recommendation pipeline to add explainability to the system. This process also helped to identify bugs and bottlenecks in the pipeline (See Figure D.3 for an example).

4 Results

This chapter outlines the computational and user study results arising from the methodology outlined in Chapter 3. Each of the dependent variables in Table 19 are comprehensively analyzed using statistical analyses to determine whether the null hypotheses in Chapter 3.1.5 hold or not.

Table 19: Summary of Dependent Variables Studied from the User Study

Type	Measure
Motivation	Pre / Post
Privacy Preferences	Pre
User Experience	Post
Projected Behaviour Performance	Post

4.1 User Study

4.1.1 Demographic Information

There was gender parity among the participants with 52.05% (38/73) being male and 47.95% (35/73) female. The most frequent age range among both males and females was 20-22 years (43.84% 32/73) (See Table E.1).

Male participants had an average weight of 81.43 kilograms, height of 1.77 meters, and BMI of 25.77 kg/m^2 . Furthermore, male participants self-reported engaging in an average of 225.79 minutes of moderate to rigorous physical activity per week. According to physical activity guidelines published by the US Department of Health and Human Services in 2018 [104, 105], 34.2% (13/38) of males were classified as being sedentary, engaging in less than 150 minutes of weekly moderate to rigorous physical activity.

Female participants had an average weight of 63.13 kilograms, height of 1.62 meters, and BMI of 23.92 kg/m^2 . Following this, they engaged in an average of 263.14 minutes of moderate to rigorous weekly physical activity. 20% (7/35) of females were classified as being sedentary as well based on the previously aforementioned physical activity guidelines [104].

A majority of the participants identified as being ethnically South Asian (31.5%, 23/73) followed by White / Caucasian (15.1%, 11/73) and Non-White West Asian (10.9%, 8/73). There was little or no Japanese (0%, 0/73), Arab (1.4%, 1/73), Korean (1.4%, 1/73), or Filipino (2.7%, 2/73) representation (See Table E.2 for more information).

The participants were shown to be highly educated with most pursuing or having completed an Undergraduate (76.7% ,56/73) or Master's (17.8%, 13/73) degree (See Table E.3 for more information). Few participants came from a Vocational College (2.7%, 2/73) or PhD Program (2.7%, 2/73). Computer Science (26.0%, 19/73), Neuroscience (13.7%, 10/73), Computer Engineering (8.2%, 6/73), and Kinesiology (6.8%, 5/73) were among the most common majors of study, highlighting high technological and STEM-based proficiency.

Lastly, Table 20 summarizes the statistics of participants that were randomly assigned into conditions of the study based on gender, system type, and group with respect to their lifestyle classification. It can be seen that the number of physically active participants within certain conditions may skew the results.

Table 20: Distribution of Participants by Gender and Physical Activity Classification into Groups and Systems

Gender	System	Group	Lifestyle		Total
			Active	Sedentary	
Male	Context	Control	8	2	10
		Experimental	6	2	8
		Total	14	4	18
	Content	Control	8	3	11
		Experimental	3	6	9
		Total	11	9	20
	Total	Control	16	5	21
		Experimental	9	8	17
		Total	25	13	38
Female	Context	Control	7	2	9
		Experimental	6	2	8
		Total	13	4	17
	Content	Control	4	1	5
		Experimental	11	2	13
		Total	15	3	18
	Total	Control	11	3	14
		Experimental	17	4	21
		Total	28	7	35
Total	Context	Control	15	4	19
		Experimental	12	4	16
		Total	27	8	35
	Content	Control	12	4	16
		Experimental	14	8	22
		Total	26	12	38
	Total	Control	27	8	35
		Experimental	26	12	38
		Total	53	20	73

4.1.2 Pre-Intervention Motivation

The grand mean for pre-intervention exercise motivation was 5.589 (SD = 0.801) after reverse-coding the 7-point Likert scales negatively phrased items. The control (M = 5.627, SD = 0.717) and experimental (M = 5.555, SD = 0.879) groups had similar motivation averages irrespective of the type of recommender system as shown in Figure 22. Internal consistency for the scale was assessed using McDonald’s Omega which demonstrated good reliability for exercise enjoyment ($\omega = 0.790$).

A 2 (*Gender*) × 2 (*Group*) × 2 (*System*) between-subjects Aligned Rank Transform ANOVA (see Table 21) demonstrated that there were no significant interaction effects among the independent variables.

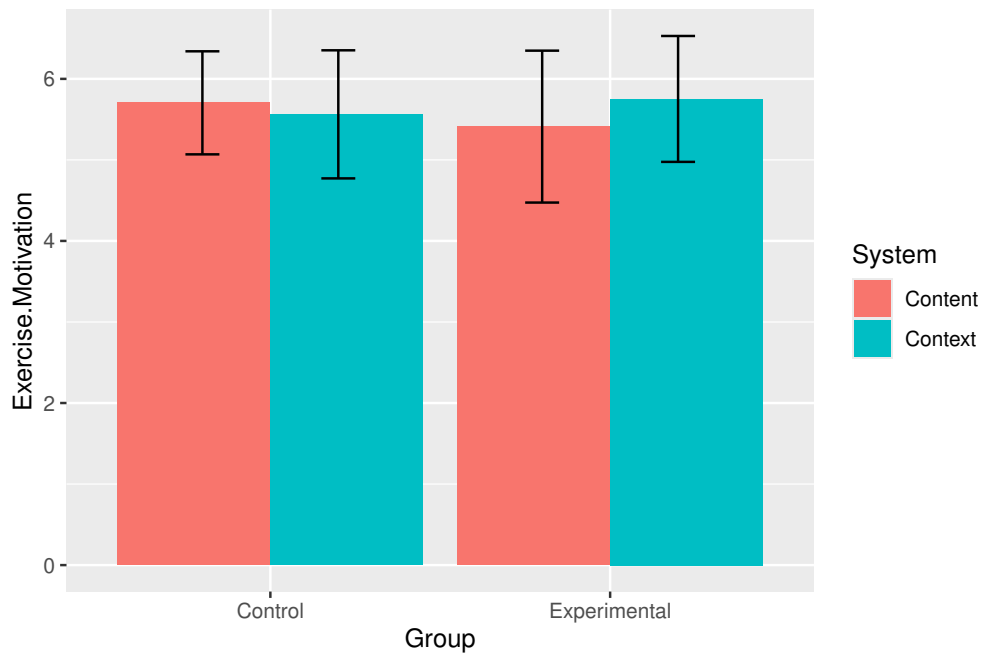


Figure 22: Mean exercise motivation scores among the study groups and recommender systems. Error bars represent $\pm 1 SD$

Table 21: Aligned Rank Transform ANOVA Results based on Pre-Intervention Motivation

	Df	Df.res	F-value	p-value
Gender	1	65	0.05	$p > 0.1$
Group	1	65	0.01	$p > 0.1$
System	1	65	0.27	$p > 0.1$
Gender:Group	1	65	0.92	$p > 0.1$
Gender:System	1	65	0.65	$p > 0.1$
Group:System	1	65	1.44	$p > 0.1$
Gender:Group:System	1	65	2.98	$p > 0.05$

A numerical but non-significant difference was observed between gender, group, and system factors (See Figure 23). A follow-up pairwise contrast test identified a significant ($p < 0.05$) and marginal ($p = 0.06$) effect, as seen in Table 22. Altogether, the results indicated that randomizing participants into the study group and recommender system type using two random number generators worked effectively.

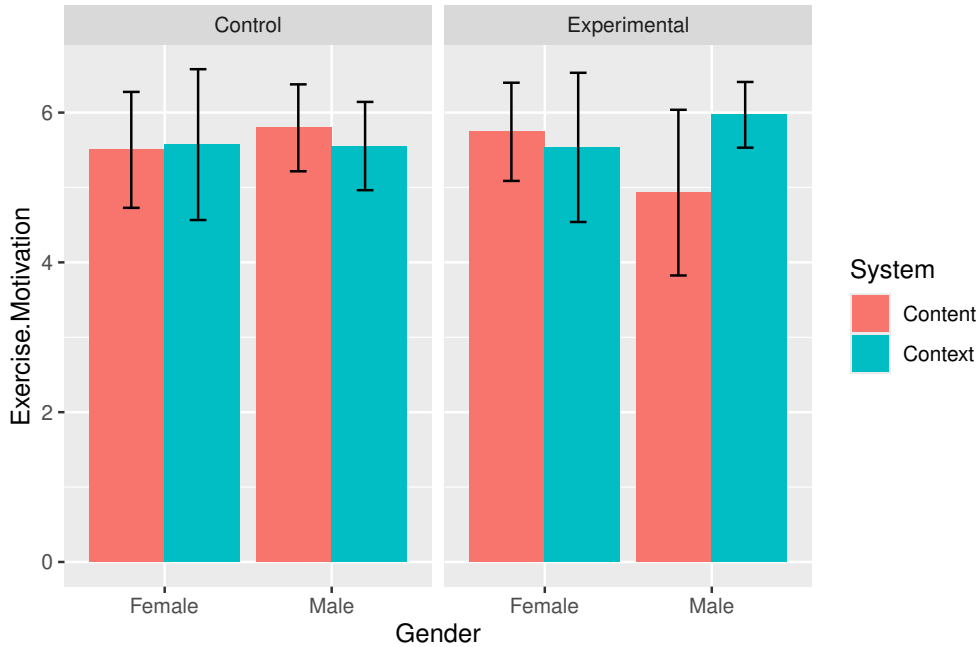


Figure 23: Mean exercise motivation scores among the $Gender \times Group \times System$ interactions. Error bars represent $\pm 1 SD$

Table 22: Pairwise Contrast Test for $Gender \times Group \times System$ from Table 21. Contrasts with $p \geq 0.1$ have been filtered out for brevity.

Contrast	Estimate	SE	df	t-ratio	p-value
Male,Experimental,Content - Male,Experimental,Context	-22.46	10.32	65	-2.17	$p < 0.05$
Male,Control,Content - Male,Experimental,Content	18.24	9.55	65	1.91	$p > 0.05$
Female,Experimental,Content - Male,Experimental,Content	15.95	9.21	65	1.73	$p > 0.05$

4.1.3 Pre-Intervention Privacy Preferences

The grand mean for privacy attitudes was 3.287 ($SD = 0.381$) after reverse-coding the 5-point Likert scales negatively phrased items. This indicated that participants were moderately privacy-conscious when using health-based recommender systems. The control ($M = 3.304$, $SD = 0.338$) and experimental ($M = 3.271$, $SD = 0.421$) groups had very similar privacy attitudes irrespective of the type of recommender system or gender differentiation (See Figures 24 and 25). Internal consistency for the scale was assessed using McDonald's Omega which demonstrated poor reliability for privacy attitudes ($\omega = 0.334$).

Assessing for an interaction effect through a 2 (*Gender*) × 2 (*Group*) × 2 (*System*) between-subjects Aligned Rank Transform ANOVA yielded no significant effects, as seen in Table 23.

Table 23: Aligned Rank Transform ANOVA Results based on Pre-Intervention Privacy Attitudes

	Df	Df.res	F-value	<i>p</i>-value
Gender	1	65	0.38	$p > 0.1$
Group	1	65	0.11	$p > 0.1$
System	1	65	2.58	$p > 0.1$
Gender:Group	1	65	0.29	$p > 0.1$
Gender:System	1	65	1.02	$p > 0.1$
Group:System	1	65	0.45	$p > 0.1$
Gender:Group:System	1	65	3.27	$p > 0.05$

There was a numerical but non-significant difference based on the gender, group, and system factors as seen in Figure 25. Pairwise contrast tests revealed two significant interactions ($p < 0.05$) and two marginal effects ($p \leq 0.054$) as seen in Table 24.

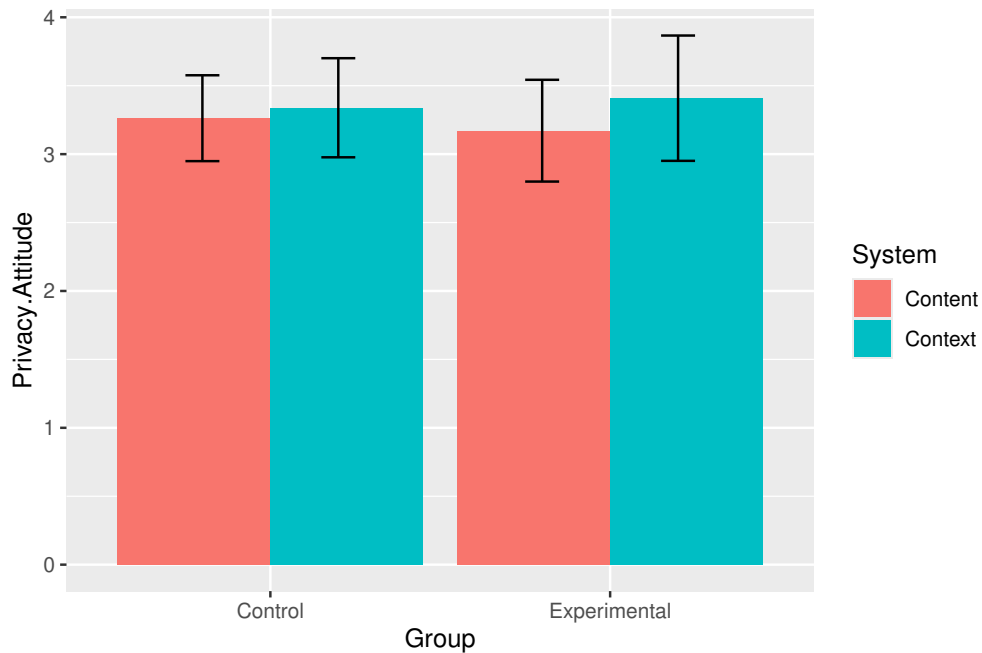


Figure 24: Mean privacy attitude scores among the study groups and recommender systems. Error bars represent $\pm 1 SD$

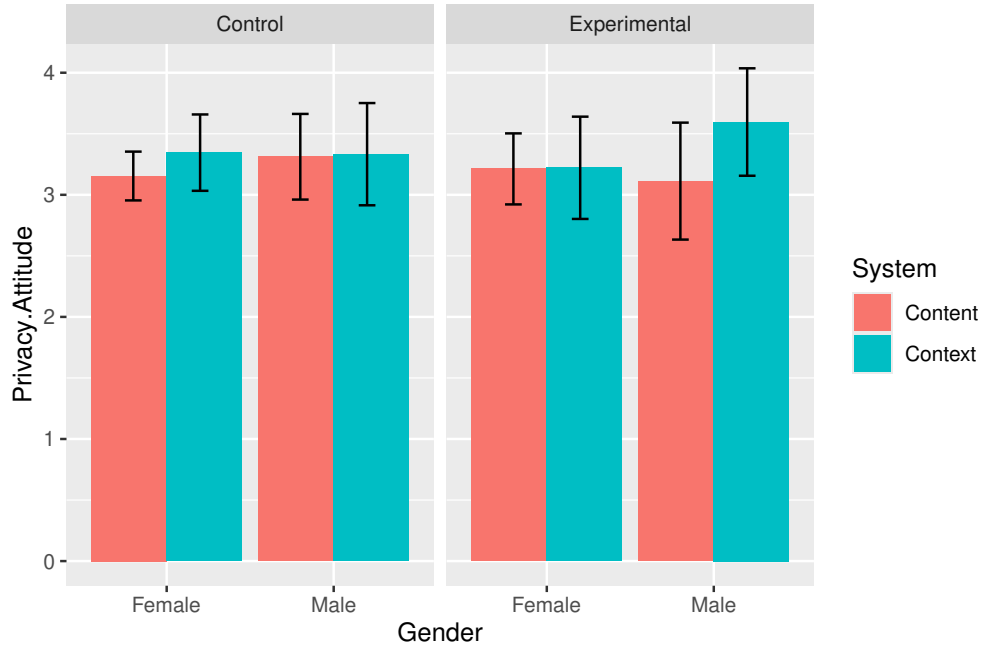


Figure 25: Mean privacy attitude scores among the $Gender \times Group \times System$ interactions. Error bars represent $\pm 1 SD$

Table 24: Pairwise Contrast Test for $Gender \times Group \times System$ from Table 23. Contrasts with $p \geq 0.1$ have been filtered out for brevity.

Contrast	Estimate	SE	df	t-ratio	p-value
Male,Experimental,Content - Male,Experimental,Context	-22.41	10.22	65	-2.19	$p < 0.05$
Female,Experimental,Context - Male,Experimental,Context	-21.44	10.51	65	-2.04	$p < 0.05$
Female,Control,Content - Male,Experimental,Context	-23.69	11.99	65	-1.98	$p > 0.05$
Female,Experimental,Content - Male,Experimental,Context	-18.53	9.44	65	-1.96	$p > 0.05$

4.1.4 User Experience

The grand mean for user experience was 3.660 ($SD = 0.883$) after reverse-coding the 5-point Likert scales negatively phrased items which was calculated by averaging each of the adapted HUED-CUP frameworks constructs from the study questionnaire. This indicated that the interventions were user-friendly and persuasive, surpassing neutral baselines (3), as seen in Figure 26. Internal consistency for the scale was assessed using McDonald's Omega which demonstrated good reliability for perceived user experience ($\omega = 0.840$). The individual constructs could not meaningfully be measured as they each consisted of 2-3 responses.

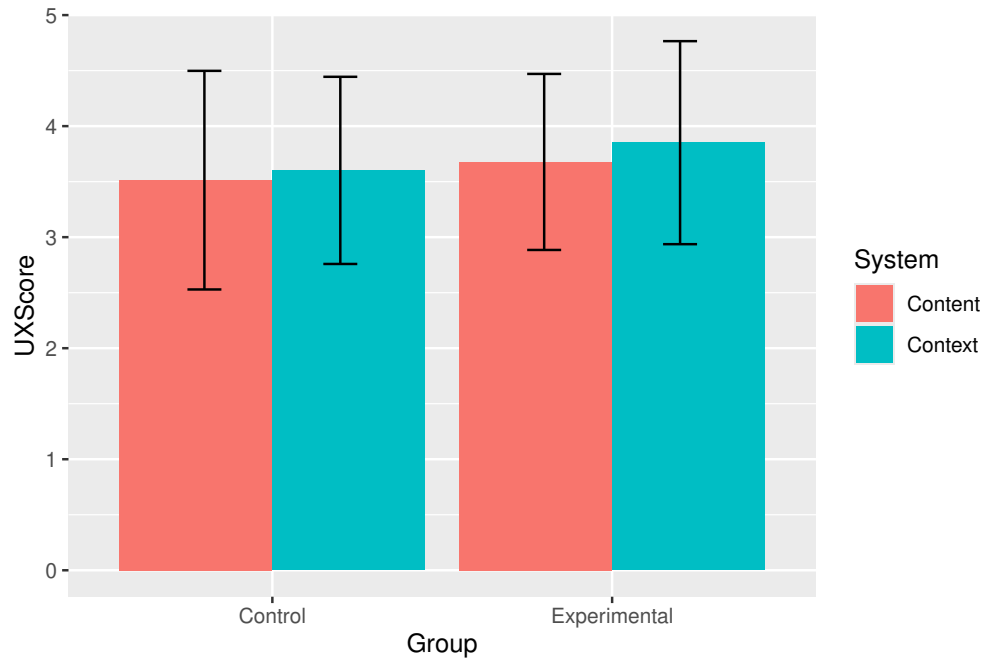


Figure 26: Mean User Experience scores among the study groups and recommender systems. Error bars represent $\pm 1 SD$

The experimental group ($M = 3.751$, $SD = 0.849$) performed the best followed by the control group ($M = 3.561$, $SD = 0.909$), however, there was no significant group effect. Furthermore, it can be seen that the context-aware recommender system was better perceived compared to the content-based recommender system across groups.

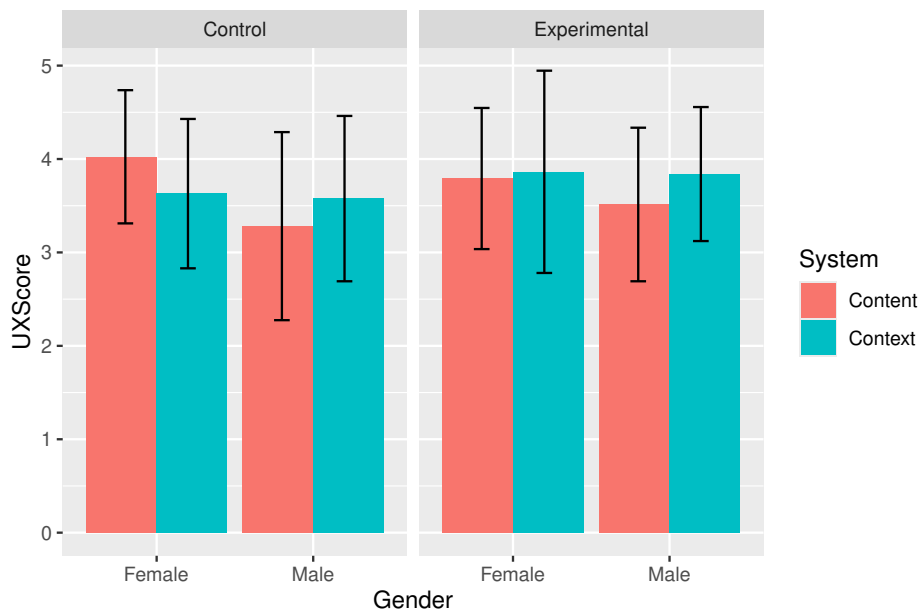


Figure 27: Mean User Experience scores among the $Gender \times Group \times System$ interactions. Error bars represent $\pm 1 SD$

A 2 (*Gender*) \times 2 (*Group*) \times 2 (*System*) \times 7 (*UX.Factor*) mixed-methods Aligned Rank Transform ANOVA (See Table 25) demonstrated that there were two significant interaction effects with relation to the type of afforded user experience (*UX.Level*) ($p < 0.001$) and how it affects the system ($p < 0.01$). Following this, there was a numerical but non-significant gender effect ($p \leq 0.07$).

Table 25: Aligned Rank Transform ANOVA Results based on Perceived User Experience

	Df	Df.res	F-value	p-value
Gender	3.46	1	65	$p > 0.05$
Group	2.21	1	65	$p > 0.1$
System	0.60	1	65	$p > 0.1$
UX.Factor	8.18	6	390	$p < 0.001$
Gender:Group	0.34	1	65	$p > 0.1$
Gender:System	1.86	1	65	$p > 0.1$
Group:System	0.13	1	65	$p > 0.1$
Gender:UX.Factor	0.57	6	390	$p > 0.1$
Group:UX.Factor	0.16	6	390	$p > 0.1$
System:UX.Factor	3.32	6	390	$p < 0.01$
Gender:Group:System	0.52	1	65	$p > 0.1$
Gender:Group:UX.Factor	0.66	6	390	$p > 0.1$
Gender:System:UX.Factor	1.61	6	390	$p > 0.1$
Group:System:UX.Factor	0.63	6	390	$p > 0.1$
Gender:Group:System:UX.Factor	0.52	6	390	$p > 0.1$

Figure 28 visualizes the differences in impact that each type of user experience factor (*UX.Level*) contributed to the overall perceived user experience afforded by the recommender systems. Usability ($M = 4.151$, $SD = 0.799$) played the largest role in a positive user experience followed by privacy ($M = 3.774$, $SD = 0.842$) and expertise ($M = 3.644$, $SD = 0.709$). Pairwise contrast testing validated the significant differences between usability and privacy as demonstrated in Table 26.

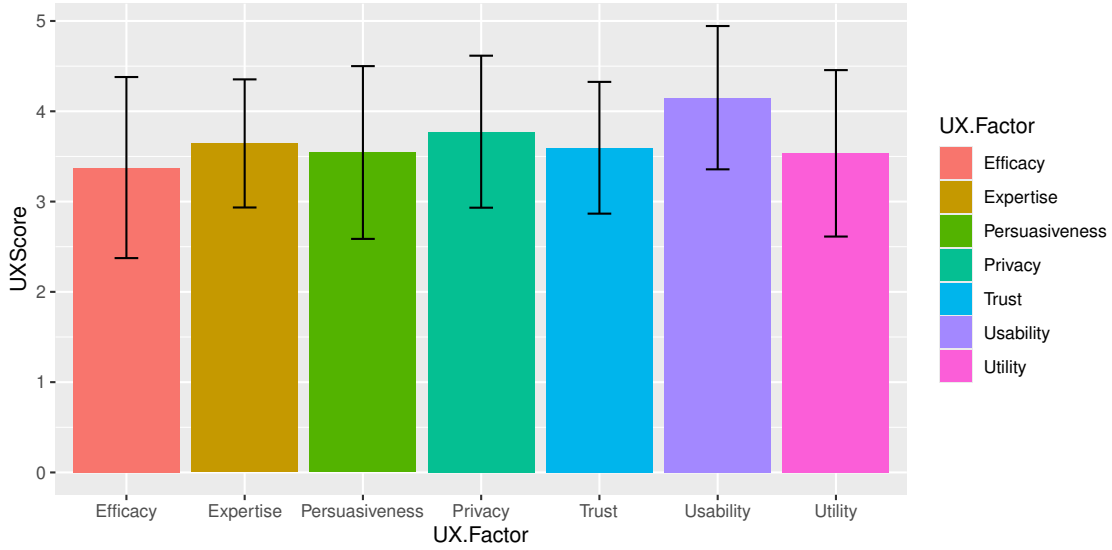


Figure 28: Mean User Experience scores based on the varying ComTech Factors. Error bars represent $\pm 1 SD$

Table 26: Pairwise Contrast Test for the User Experience Factor from Table 25. Contrasts with $p \geq 0.1$ have been filtered out for brevity.

Contrast	Estimate	SE	df	t-ratio	p-value
Efficacy - Usability	-131.77	21.44	390	-6.15	$p < 0.001$
Trust - Usability	-111.06	21.44	390	-5.18	$p < 0.001$
Persuasiveness - Usability	-108.36	21.44	390	-5.05	$p < 0.001$
Usability - Utility	105.82	21.44	390	4.94	$p < 0.001$
Expertise - Usability	-98.43	21.44	390	-4.59	$p < 0.001$
Privacy - Usability	-71.64	21.44	390	-3.34	$p < 0.001$
Efficacy - Privacy	-60.13	21.44	390	-2.80	$p < 0.05$
Privacy - Trust	39.42	21.44	390	1.84	$p > 0.05$
Persuasiveness - Privacy	-36.72	21.44	390	-1.71	$p > 0.05$

Figure 29 visualizes how the user experience factors differ based on the type of system. Figure 30 also shows how user experience differs across different subgroups based on the gender, group, system type, and afforded user experience factor. In both cases, the general trends are consistent.

Pairwise contrast testing provided in Table 27 again confirmed that usability and privacy were significantly greater in the context-aware condition rather than the content-based recommender system.

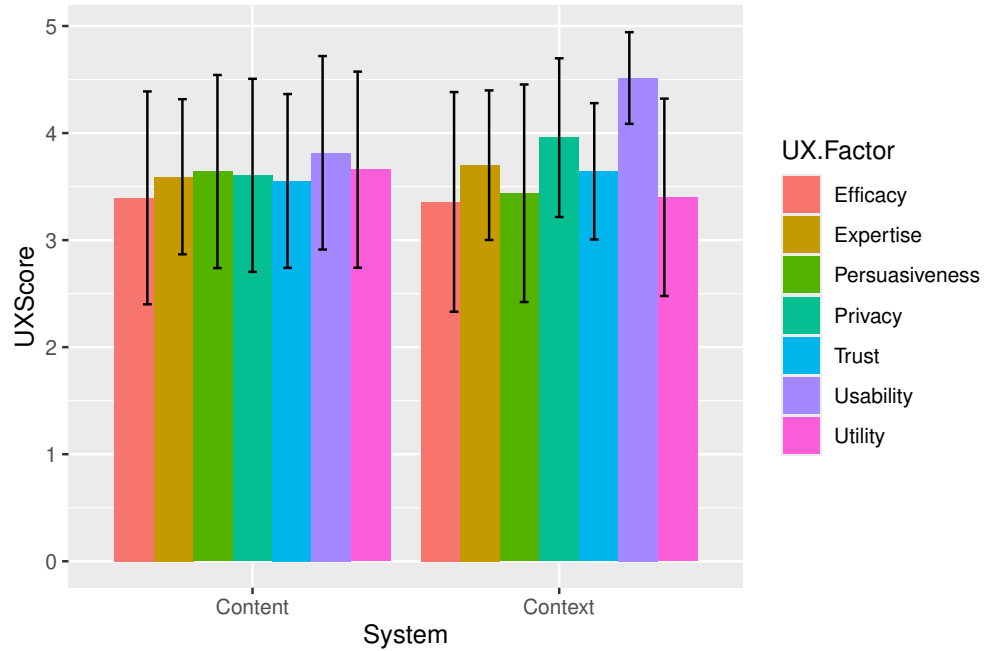


Figure 29: Mean User Experience scores based on the $System \times UX.Factor$ interactions. Error bars represent ± 1 SD

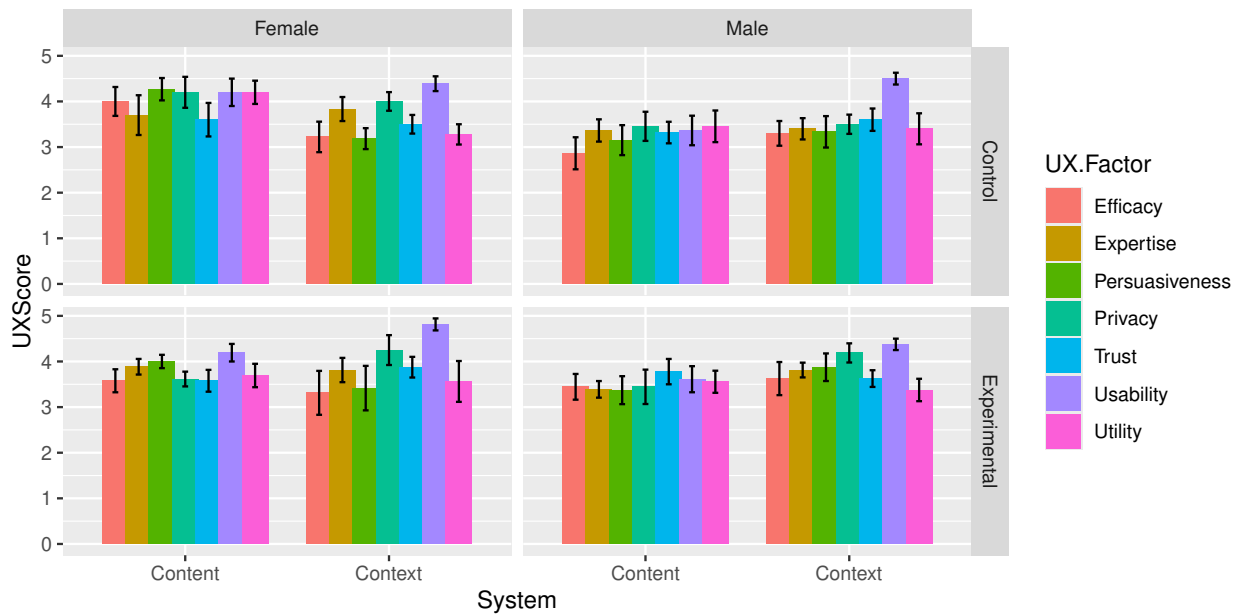


Figure 30: Mean User Experience Scores based on Subgroups: Gender, System, and UX.Factor. Error bars represent ± 1 SD

Table 27: Pairwise Contrast Test for $System \times UX.Factor$ from Table 25. Contrasts with $p \geq 0.1$ have been filtered out for brevity.

Contrast	Estimate	SE	df	t-ratio	p-value
Context,Usability - Context,Utility	198.21	29.40	390.00	6.74	$p < 0.001$
Context,Efficacy - Context,Usability	-195.49	29.40	390.00	-6.65	$p < 0.001$
Context,Persuasiveness - Context,Usability	-189.31	29.40	390.00	-6.44	$p < 0.001$
Content,Efficacy - Context,Usability	-206.94	35.65	290.64	-5.80	$p < 0.001$
Context,Trust - Context,Usability	-168.52	29.40	390.00	-5.73	$p < 0.001$
Context,Expertise - Context,Usability	-160.43	29.40	390.00	-5.46	$p < 0.001$
Content,Trust - Context,Usability	-185.75	35.65	290.64	-5.21	$p < 0.001$
Content,Expertise - Context,Usability	-176.42	35.65	290.64	-4.95	$p < 0.001$
Content,Privacy - Context,Usability	-166.51	35.65	290.64	-4.67	$p < 0.001$
Content,Persuasiveness - Context,Usability	-166.42	35.65	290.64	-4.67	$p < 0.001$
Content,Utility - Context,Usability	-159.47	35.65	290.64	-4.47	$p < 0.001$
Content,Usability - Context,Usability	-137.94	35.65	290.64	-3.87	$p < 0.001$
Context,Privacy - Context,Usability	-110.54	29.40	390.00	-3.76	$p < 0.001$
Context,Privacy - Context,Utility	87.67	29.40	390.00	2.98	$p < 0.05$
Context,Efficacy - Context,Privacy	-84.96	29.40	390.00	-2.89	$p < 0.05$
Content,Efficacy - Context,Privacy	-96.40	35.65	290.64	-2.70	$p < 0.05$
Context,Persuasiveness - Context,Privacy	-78.77	29.40	390.00	-2.68	$p < 0.05$
Content,Efficacy - Content,Usability	-68.99	29.96	390.00	-2.30	$p < 0.05$
Content,Trust - Context,Privacy	-75.21	35.65	290.64	-2.11	$p < 0.05$
Context,Privacy - Context,Trust	57.98	29.40	390.00	1.97	$p < 0.05$
Content,Expertise - Context,Privacy	-65.88	35.65	290.64	-1.85	$p > 0.05$
Context,Expertise - Context,Privacy	-49.89	29.40	390.00	-1.70	$p > 0.05$
Content,Usability - Context,Utility	60.26	35.65	290.64	1.69	$p > 0.05$

Altogether, the results indicate that while usability and privacy were important to designing an effective recommender system, the numerical differences were not enough to create a significant difference between the experimental and control group. Careful consideration is required to understand how to best select and apply persuasive design techniques to fitness-based recommender systems.

4.1.5 Perceived Post-Intervention Motivation

The grand mean for post-intervention exercise motivation was 0.546 (SD = 0.188) after taking the normalized [106] averages of self-efficacy resulting from a 7-point Likert scale and self-regulation from two 5-point Likert scales. There was no reverse-coding associated with any of the items. Internal consistency for the scales was assessed using McDonald's Omega for self-efficacy ($\omega = 0.767$) and self-regulation ($\omega = 0.656$) indicating reasonable reliability. Figure 31 shows the post-intervention exercise motivation scores across each group and type of system. While the control (M = 0.55, SD = 0.206) and experimental (M = 0.540, SD

= 0.172) groups had similar means, the experimental context-aware recommender system ($M = 0.583$, $SD = 0.151$) performed the best followed by the control conditions content-based ($M = 0.573$, $SD = 0.186$) and context-aware systems ($M = 0.533$, $SD = 0.222$). The experimental content-based system ($M = 0.509$, $SD = 0.181$) performed the worst.

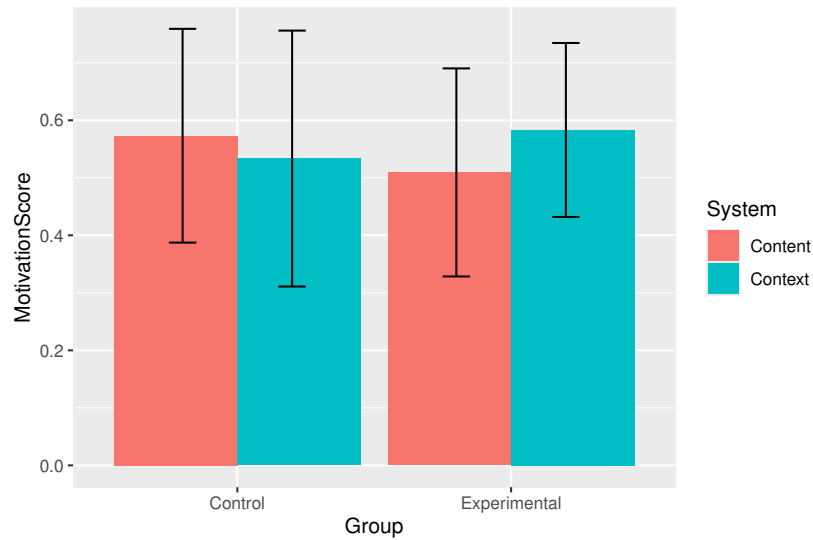


Figure 31: Mean Post-Intervention Exercise Motivation scores among the study groups and recommender systems. Error bars represent ± 1 SD

When taking gender into account between the groups and system types, Figure 32 shows small but subtle variances. Males had higher post-intervention exercise motivation scores in the control group compared to the experimental group. Similarly, females had marginally higher exercise scores in the experimental group compared to the control group.

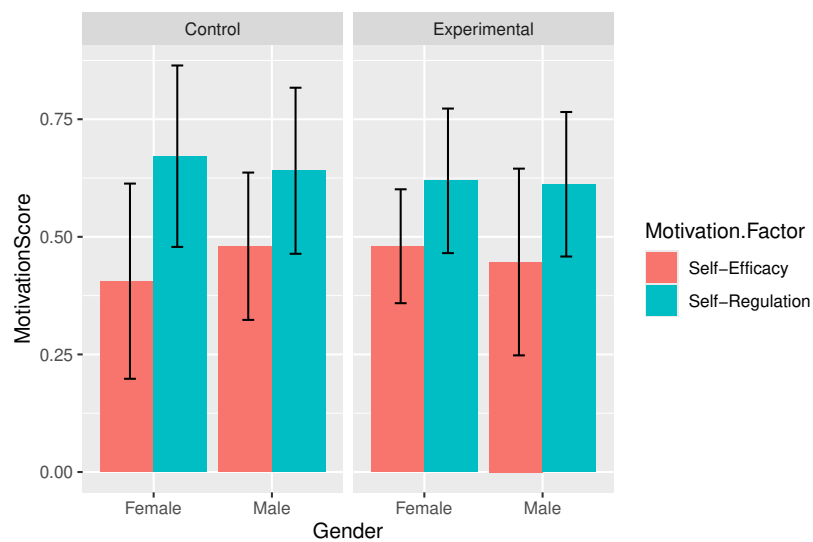


Figure 32: Mean Post-Intervention Exercise Motivation scores based on the $Gender \times Group \times System$ interactions. Error bars represent ± 1 SD

A 2 (*Gender*) × 2 (*Group*) × 2 (*System*) × 2 (*Motivation.Factor*) mixed-methods Aligned Rank Transform ANOVA (See Table 28) demonstrated a significant interaction effect ($p < 0.001$) between motivation factors. There were also two numerical but non-significant differences between the group and system ($p > 0.05$) and gender, group, and motivation factors ($p > 0.05$).

Table 28: Aligned Rank Transform ANOVA Results based on Perceived Post-Intervention Exercise Motivation

	Df	Df.res	F-value	p-value
Gender	0.05	1	65	$p > 0.1$
Group	0.12	1	65	$p > 0.1$
System	0.42	1	65	$p > 0.1$
Motivation.Factor	68.23	1	65	$p < 0.001$
Gender:Group	0.84	1	65	$p > 0.1$
Gender:System	0.06	1	65	$p > 0.1$
Group:System	3.18	1	65	$p > 0.05$
Gender:Motivation.Factor	0.64	1	65	$p > 0.1$
Group:Motivation.Factor	1.87	1	65	$p > 0.1$
System:Motivation.Factor	0.15	1	65	$p > 0.1$
Gender:Group:System	0.32	1	65	$p > 0.1$
Gender:Group:Motivation.Factor	2.90	1	65	$p > 0.05$
Gender:System:Motivation.Factor	0.41	1	65	$p > 0.1$
Group:System:Motivation.Factor	0.01	1	65	$p > 0.1$
Gender:Group:System:Motivation.Factor	1.36	1	65	$p > 0.1$

Pairwise contrast testing on the motivation factor identified a significant relationship between self-efficacy and self-regulation ($p < 0.001$). A positive correlation between the two constructs was further validated through a Spearman's Rho test ($\rho = 0.385$, $p < 0.001$); as self-regulation increases, so does self-efficacy.

Following this, another pairwise contrast test on the gender, group, and motivation factor variables uncovered findings in Table 29. The significant effects were caused due to the larger values of self-regulation compared to self-efficacy across all groups and gender conditions, however, there were no results which implied that by themselves, self-efficacy or self-regulation were significantly different across gender or groups.

Table 29: Pairwise Contrast Test for *Gender × Group × Motivation.Factor* from Table 28. Contrasts with $p \geq 0.1$ have been filtered out for brevity. Abbreviations: SE Self-efficacy, SR Self-regulation

Contrast	Estimate	SE	df	t-ratio	p-value
Female,Control,SE - Female,Control,SR	-63.87	11.73	65.00	-5.45	$p < 0.001$
Female,Control,SE - Male,Control,SR	-56.71	13.61	111.99	-4.17	$p < 0.001$
Male,Control,SE - Male,Control,SR	-36.75	9.18	65.00	-4.00	$p < 0.001$
Female,Control,SE - Female,Experimental,SR	-52.55	13.76	111.99	-3.82	$p < 0.001$
Male,Experimental,SE - Male,Experimental,SR	-38.38	10.21	65.00	-3.76	$p < 0.001$
Female,Control,SR - Male,Experimental,SE	51.73	14.21	111.99	3.64	$p < 0.001$
Female,Experimental,SE - Female,Experimental,SR	-35.11	9.45	65.00	-3.72	$p < 0.001$
Female,Control,SE - Male,Experimental,SR	-50.52	14.21	111.99	-3.56	$p < 0.001$
Male,Control,SR - Male,Experimental,SE	44.57	12.55	111.99	3.55	$p < 0.001$
Female,Control,SR - Female,Experimental,SE	46.42	13.76	111.99	3.37	$p < 0.01$
Female,Experimental,SE - Male,Control,SR	-39.26	12.04	111.99	-3.26	$p < 0.01$
Female,Control,SR - Male,Control,SE	43.91	13.61	111.99	3.23	$p < 0.01$
Female,Experimental,SR - Male,Experimental,SE	40.42	12.71	111.99	3.18	$p < 0.01$
Female,Experimental,SR - Male,Control,SE	32.59	12.04	111.99	2.71	$p < 0.01$
Female,Experimental,SE - Male,Experimental,SR	-33.07	12.71	111.99	-2.60	$p < 0.05$
Male,Control,SE - Male,Experimental,SR	-30.56	12.55	111.99	-2.43	$p < 0.05$

In conclusion, the lack of a meaningful significant effect may have been due small group sizes or too little emphasis on the motivational factors that contributed to the experimental groups fitness recommendations during the system design phase. Pairwise contrast testing on the group and system revealed no significant results ($p < 0.05$).

4.1.6 Open-Ended Comments (Self-Efficacy)

Participants were asked to write about their exercise self-efficacy and provide justifications for their current behaviour. The author performed sentiment analysis with Roberta Base Sentiment [100]. Initial coding resulted in a Cohen Kappa score of 0.44 (67.0%), indicating moderate agreement [107]. After revising the discrepancies between labeling, a majority of the comments were classified as being neutral (64.4%, 47/73) followed by an equal amount of positive (17.8%, 13/73) and negative (17.8%, 13/73) sentiments.

A Chi-Square test was undertaken to evaluate the relationship between the gender, group, and system variables with respect to sentiment counts. Table 30 shows that there were no significant interactions ($p < 0.05$), meaning that sentiment distributions did not largely differ between each of the groupings.

Table 30: Chi-Square Test Results for Gender, Group, and System Variables Based on Sentiment Counts of Open-Ended Self-Efficacy Responses

Variable		Value	df	p-value
Gender	χ^2	0.67	2	$p > 0.1$
	N	73		
Group	χ^2	1.69	2	$p > 0.1$
	N	73		
System	χ^2	0.67	2	$p > 0.1$
	N	73		

Thematic coding was undertaken to analyze positive and negative trends that affected exercise self-efficacy. The findings are summarized in Table 31 (Barriers) and 32 (Enablers).

Table 31: Thematic Coding: Sample Barriers to Self-Efficacy Questions

Theme	ID	Comment
Injury	S63	<i>“When it comes to physical discomfort, I tend to skip physical exercise in general because I don’t want to make that discomfort worse, however if it’s something that’s cognitive, such as an emotion I’m feeling, I tend to push through otherwise.”</i>
Weather	S62	<i>“In general I am not very consistent when it comes to working out. However, once I do start I can usually go to the gym 3-4 times a week. But I am also lazy and if the weather is gloomy or I’m feeling extra lazy then I will exercise less or just do it the next day [...]”</i>
Mood	S20	<i>“Sometimes, my exercise routine depends on my mood. Even with an exercise routine, I may not follow through on it if it doesn’t account for how I’m feeling.”</i>
Time Management	S19	<i>“[...] instead, if I have some deadlines coming, I will feel some things stay in my mind and have less determination or motivation to do exercise.”</i>

Table 31 (continued)

Theme	ID	Comment
Explicit Instruction	S17	<i>“I would like everything planned out, otherwise I would quit really fast. I started many exercises and I quit them because I didn’t plan out the exercise very well and I feel they are too complicated to continue.”</i>
Variety (Boredom)	S23	<i>“I hate exercise for exercise sake. I want to do a task and show off my progress to my friends. If it is repetitive, boring, or I come up with a fun task to do, I will switch tasks because I cant stay focused on going nowhere on repeat [...]”</i>
Menstrual Cycle	S73	<i>“I am emotion driven, and my body works on a monthly cycle, so not every day or every week will be the same.”</i>
Difficulty	S66	<i>“More repetitive exercises or really hard exercise will either make me give up or come back for modification.”</i>

Table 32: Thematic Coding: Enablers on Open-Ended Exercise Self-Efficacy Questions

Theme	ID	Comment
Consistency	S44	<i>“Consistency is the key in doing exercises and maintaining fitness. Exercise helps me to drift away and relax a bit. Aside from that it helps me create structure in my day, week, and life.”</i>
Motivation	S65	<i>“I work better when someone pushes or motivates me. It is especially important for someone like me who lives alone [...]”</i>
Progress Tracking	S14	<i>“The most important thing is progress. If I see it, I’ll continue.”</i>
Enjoyment	S61	<i>“During the winter semester, even though I took 6 classes at once, I still made time to workout (weightlifting) at least 3 times a day, even though I was frequently tired, stressed out and really busy. I enjoy weightlifting and it helps calm me which gives me an incentive to do it even if I am busy. [...]”</i>

The results demonstrate that there are many more barriers to address when trying to increase exercise motivation compared to enablers. In attempting to increase exercise motivation, both factors must thoughtfully be considered simultaneously. Overall, the identified findings and thematic codes are consistent with related work [108, 109].

4.1.7 Open-Ended Comments (Self-Regulation)

Following the open-ended questions about self-efficacy, participants were asked to write about how self-regulation affected their current exercise motivation and behaviour. Initial sentiment coding alongside Roberta Base Sentiment [100] resulted in a Cohen Kappa score of 0.61 (79.0%), indicating substantial agreement [107]. Reviewing and correcting the discrepancies between the labels resulted in a majority of neutral comments (58.9%, 43/73) followed by positive (34.25%, 25/73) and negative (6.85%, 5/73) sentiments.

A Chi-Square test was undertaken to evaluate the relationship between the gender, group, and system variables with respect to sentiment counts. Table 33 shows that there were no significant interactions ($p < 0.05$), meaning that sentiment distributions did not largely differ between each of the groupings.

Table 33: Chi-Square Test Results for Gender, Group, and System Variables Based on Sentiment Counts of Open-Ended Self-Regulation Responses

Variable		Value	df	p-value
Gender	χ^2	1.10	2	$p > 0.1$
	N	73		
Group	χ^2	1.58	2	$p > 0.1$
	N	73		
System	χ^2	0.14	2	$p > 0.1$
	N	73		

Thematic coding was undertaken to identify common strategies to self-regulate adequate levels of physical activity into a weekly schedule. The finalized results are summarized in Table 34.

Table 34: Thematic Coding: Strategies for Self-Regulating Exercise Adoption

Theme	Example (ID)	Comment
Progress Tracking	S03	<i>“I have a list of various exercises in a routine style basis. Setting goals and clear steps help to track progress and accomplish set goals. However, as it is personal journey, I don’t feel he need to boast about goals or progress to others.”</i>
Motivation	S08	<i>“[...] I am very goal and numbers driven. If I see my strength go up, this is motivating, if I see my body start to look better, this is motivating, if I feel more mobile and agile, this is motivating, if I am better at a sport I do, this is motivating [...]”</i>
Socializing	S14	<i>“I think being in touch with others is the key ingredient for me. Gym buddies and what not.”</i>

Table 34 (continued)

Theme	Example (ID)	Comment
Spontaneity	S19	<i>“I prefer not to have predefined or fixed schedule or plan for exercise. Instead, I prefer a flexible plan that I will try to accomplish.”</i>
Explicit Instruction	S23	<i>“I need schedules and steps to get things done but I struggle to remember to use them.”</i>
Goal-Setting	S49	<i>“I do not give up until I achieve my goal. Exercise and being fit is one of these goals.”</i>
Consistency	S70	<i>“My answers came from personal reflection as well as previous experiences I have had planning my workouts and achieving fitness or performance goals. My biggest struggle is the amount of additional activities I have on a weekly basis, such as attending university full time, having 2 part time jobs, being married, etc, which leave me with little to no room for free time to use in my workouts.”</i>

In assessing the results, it is evident that self-efficacy and self-regulation are closely aligned with each other. Many strategies to regulate exercise stemmed from using enabling self-efficacy themes highlighted in Table 32.

4.1.8 Projected Post-Intervention Behaviour Performance

Prior to assessing the results, two participants data were removed from analyses as their self-reported physical activity per week or new personal goal (in minutes) were outliers (> 600).

The grand mean for exercise score (projected post-intervention behaviour performance) was 224.93 minutes of physical activity per week (SD = 127.352). The control group reported slightly higher exercise scores (M = 237.754, SD = 135.915) compared to the experimental group (M = 217.733, SD = 119.400). Figure 33 shows the exercise scores across each group and type of system.

The control groups context-aware system performed best (M = 261.554, SD = 150.080) followed by its experimental counterpart (M = 258.750, SD = 120.423) and the control groups content-based system (M = 199.453, SD = 110.588). Lastly, the experimental content-based system was the worst at positively impacting projected behaviour performance (M = 199.453, SD = 110.588).

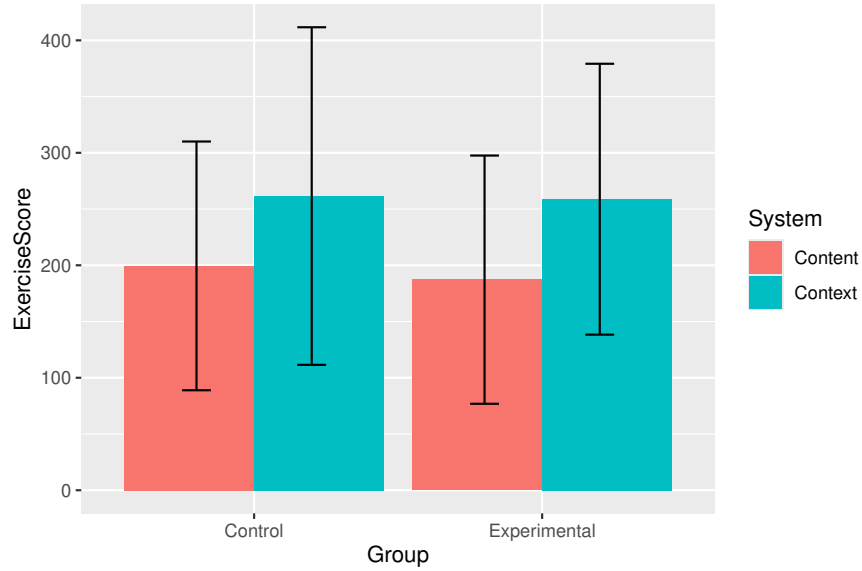


Figure 33: Mean Post-Intervention Exercise scores among the study groups and recommender systems. Error bars represent $\pm 1 SD$

Taking gender into consideration on the group and system type as seen in Figure 34 also revealed similar trends. Regardless of the gender or group, the context-aware system was able to consistently generate higher levels of positive projected behaviour performance.

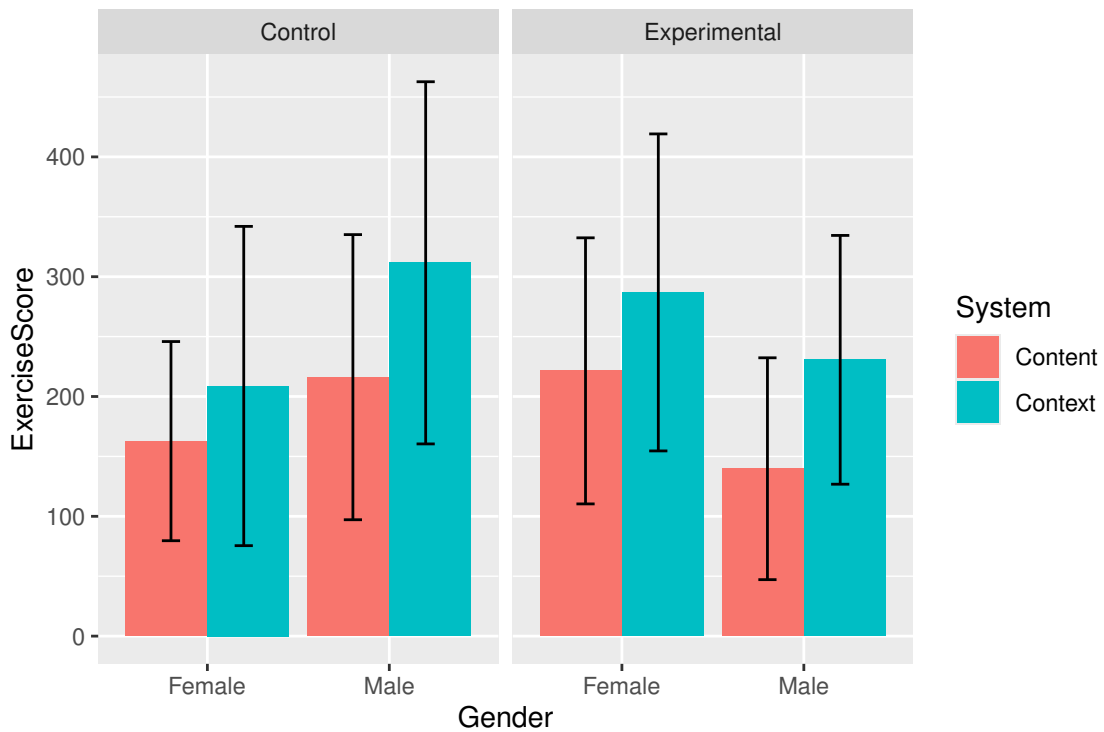


Figure 34: Mean Post-Intervention Exercise scores based on the $Gender \times Group \times System$ interactions. Error bars represent $\pm 1 SD$

A 2 (*Gender*) × 2 (*Group*) × 2 (*System*) × 2 (*Intervention*) mixed-methods Aligned Rank Transform ANOVA (See Table 35) revealed two significant interaction effects ($p < 0.01$) based on the system, gender, group, and intervention factors.

Pairwise contrast testing on the system type confirmed significant differences between the effectiveness of the context-aware and content-based recommender systems ($p < 0.01$). The context-aware system was significantly better at positively impacting projected post-intervention behaviour performance compared to the content-based system.

Additionally, pairwise contrast testing on the gender and group combinations revealed significant ($p < 0.05$) differences reported in Table 36. Firstly, males in the control group reported significantly higher projected behaviour performance scores compared to the experimental group. Following this, females in the experimental group also reported significantly higher projected behaviour performance scores compared to males in the experimental group.

Table 35: Aligned Rank Transform ANOVA Results based on Projected Post-Intervention Behaviour Performance Assessment

	Df	Df.res	F-value	p-value
Gender	0.10	1	64.59	$p > 0.1$
Group	0.11	1	64.60	$p > 0.1$
System	7.06	1	64.60	$p < 0.01$
Intervention	0.61	1	63.53	$p > 0.1$
Gender:Group	6.96	1	64.60	$p < 0.05$
Gender:System	0.84	1	64.60	$p > 0.1$
Group:System	0.33	1	64.59	$p > 0.1$
Gender:Intervention	1.60	1	63.54	$p > 0.1$
Group:Intervention	0.75	1	63.55	$p > 0.1$
System:Intervention	0.24	1	63.52	$p > 0.1$
Gender:Group:System	0.26	1	64.60	$p > 0.1$
Gender:Group:Intervention	0.01	1	63.55	$p > 0.1$
Gender:System:Intervention	0.03	1	63.54	$p > 0.1$
Group:System:Intervention	0.86	1	63.54	$p > 0.1$
Gender:Group:System:Intervention	2.65	1	63.53	$p > 0.1$

Table 36: Pairwise Contrast Test for *Gender* × *Group* from Table 35. Contrasts with $p \geq 0.1$ have been filtered out for brevity.

Contrast	Estimate	SE	df	t-ratio	p-value
Male,Control - Male,Experimental	26.04	12.01	64.77	2.17	$p < 0.05$
Female,Experimental - Male,Experimental	25.76	12.15	64.51	2.12	$p < 0.05$

4.1.9 Findings

4.1.10 Explainability (H1)

Analysis revealed that adding explainable metrics and visualizations to the recommender system improved user experience compared to the control group, however, the group effect was not statistically significant. As a result, H_1 had to be rejected. It may be worth investigating different types of explainable user interface modalities (e.g., Rule-based Reasoning [110]) as participant feedback suggested that the SHAP and LIME graphs were too confusing to interpret without explicit elaboration (See Chapter 4.1.4). Feedback from a participant (S71) who used the experimental content-based system also substantiated the theory that SHAP and LIME graphs are insufficient at adding explainability to the system in isolation as demonstrated by their feedback, “[...] *In terms of communication, I don’t think that users need to know about how their inputs affected the generation of their results. This is more for the engineers and software engineers that would care. It detracts from the message to a layperson.*”

4.1.11 Psychosocial Frameworks (H2, H3)

Firstly, analysis revealed that the application of psychosocial theory-driven frameworks used to design the experimental groups recommender systems did not have a significant effect on projected post-intervention behaviour performance. The control group had marginally higher exercise scores compared to the experimental group. As a result, H_2 had to be rejected. Pairwise contrast tests on the significant gender and group combinations revealed how females in the experimental group reported significantly higher projected behaviour performance scores compared to males. Additionally, pairwise contrast testing on the non-significant gender, group, and motivation factors combinations only revealed a difference in higher levels of self-regulation compared to self-efficacy across all groups and genders.

Following this, a similar analysis revealed that the application of psychosocial frameworks used to design the experimental groups recommender systems did not have a significant effect on post-intervention exercise motivation. The control and experimental groups had nearly identical averages in motivation scores. As a result, H_3 had to be rejected. A Spearman’s Rho test based on exercise motivation and projected behaviour performance demonstrated a significant positive correlation which also helps to explain the non-significant group effects ($\rho = 0.353$, $p < 0.001$).

4.1.12 User Experience (H4)

Analysis revealed that increased user experience through the adoption of explainable and persuasive design elements in the experimental groups recommender system design did not significantly effect projected post-intervention behaviour performance. The control and experimental groups had nearly identical averages.

A correlation analysis using Spearman’s Rho to determine positive relationships between user experience and post-intervention exercise motivation yielded non-significant results ($\rho = 0.126$, $p > 0.1$). A similar correlation analysis was also revealed no relationship between user experience and projected post-intervention behaviour performance ($\rho = 0.035$, $p > 0.1$). Altogether, the results indicate that that H_4 had to be rejected.

4.2 Computational Results

4.2.1 Content-based Recommender System

Table 37 shows the results of the seven evaluated machine learning models on Murarka’s Gym Exercise Dataset [3] using a 60% training, 20% testing, and 20% validation split with weighted F1-score as the evaluation metric. Table 38 outlines the performance of each model on the testing set using sci-kit learn’s classification report with weighted averages. Precision, Recall, and F1-score were used as the evaluation metrics.

Overall, it can be seen that each of the models may be overfitting to the training data. While the models generally performed equally across the evaluations, the Decision Tree had a slight advantage whereas the Perceptron performed the worst. The variance in fitting time between the models could be explained due to more complex models requiring further computations for their learning methods (e.g., backpropagation) or having to deal with high cardinality resulting from one-hot encoding.

Table 37: Weighted F1-Score on Training, Testing, and Validation Splits for the Content-Based Recommender System

Machine Learning Model	Training Set	Testing Set	Validation Set
Random Forest	0.89	0.63	0.60
Decision Tree	0.89	0.69	0.62
Logistic Regression	0.87	0.66	0.59
Gradient Boosting Trees	0.88	0.66	0.59
Multilayer Perceptron	0.88	0.64	0.63
Perceptron	0.77	0.59	0.49
XGBoost	0.88	0.65	0.59

Table 38: Classification Report Results on Testing Split for the Content-Based Recommender System

Machine Learning Model	Precision	Recall	F1-Score	Fitting Time
Random Forest	0.62	0.65	0.63	< 1 sec
Decision Tree	0.69	0.69	0.60	< 1 sec
Logistic Regression	0.59	0.62	0.60	1 sec
Gradient Boosting Trees	0.65	0.67	0.67	7 secs
Multilayer Perceptron	0.68	0.68	0.68	1 sec
Perceptron	0.57	0.64	0.59	< 1 sec
XGBoost	0.64	0.66	0.65	< 1 sec

5 Discussion

This study demonstrates that context-aware recommender systems offer better performance to the content-based alternative in all measures of the dependent variables.

Using the content-based recommender system required explicit domain knowledge of muscle groups and gym equipment whereas the context-aware condition leveraged very intuitive and well-known features such as age, gender, weight, and height. From the open-ended feedback, many participants explicitly requested visual aids or filtering options for the content-based system to select target and secondary muscles. Table 39 outlines direct feedback from participants to support the differences in usability between the content-based and context-aware systems that may have significantly impacted perceived user experience.

Table 39: Subset of Open-ended feedback on the usability of the recommender system(s)

ID	Group	System	Comment	Feedback
S06	Experimental	Content-based	<i>“I like how it was organized into telling me why the recommendation helps, how I should incorporate it into my schedule, and what my schedule should be. However, parts of the choices were tricky to understand, especially for a non-athlete like me, but luckily the experiment conductor was good at explaining whatever details I might not have known yet.”</i>	<i>“Try to add explanations about various terms since not everyone may be familiar with them and there won’t always be a person to explain them.”</i>
S11	Control	Content-based	<i>“The anatomical terms for the body parts confused me as someone who doesn’t know anatomy that well. I feel as if a diagram and a corresponding muscle group something visual would help.”</i>	<i>“With each muscle group, give examples of sports or daily tasks that would benefit from that muscle group being exercised.”</i>

Table 39 (continued)

ID	Group	System	Comment	Feedback
S26	Experimental	Context-Aware	<i>“I liked the simple-to-use interface with clear instructions that detailed all the required information as well as how it would be used to generate recommendations. What worked well was the direct use of my personal information to give me subjective recommendations that suit me. I liked the brief overview and rationale that was given based on my information before moving on to the actual recommendations. I also liked how the recommender system provided recommendations for not only exercises, but food and diet options, and ways to hold myself accountable when implementing the recommendations.”</i>	<i>“I think the fitness-based recommender system could be improved by considering the user’s current exercise habits and exercise goals. Knowing current exercise habits (e.g. how frequently the user exercises every week, how long they typically exercise for) as well as their goals (e.g. whether they want to gain muscle, lose fat or improve their cardiovascular health) would help to generate recommendations that are more tailored to their current lifestyle and future goals.”</i>

Table 39 (continued)

ID	Group	System	Comment	Feedback
S04	Experimental	Context-Aware	<i>“It was simple to use and to read. So I think for a beginner it would be helpful and easy to use. Personally, I am someone who likes numbers and concrete examples. It would have been nice if they broke down the numbers of macros to eat each day. Recommended exercises with the number of sets and reps to do. How many exercises to do in a day. The recommendations were helpful but just broad.”</i>	<i>“I think that the suggestions are great for what they address with sedentary behaviour. The recommendations are very small and easy changes that people can make in their lives. I think that having some more concrete numbers would be helpful as people can see what numbers they should aim for. However, I understand that this may be triggering to some people so it is a fine line.”</i>

When considering gender differences, the same trends were also observed in Figure 27, however, females tended to show a slight preference for the control groups content-based system instead of the experimental alternative (Figure 27). This could likely be due to the biased participant demographics as explained in Chapter 4.1.1. Participants with existing fitness plans may have not needed the comprehensive explanations and persuasive strategies afforded by the experimental group. For example, a participant (S43) in the experimental group expressed their dislike of persuasive elements embedded in the fitness plan: *“[...] Finally, I noticed that the recommendations were separated in different categories (hope, pleasure, etc.). **Maybe there could be a question in the beginning to know what advice they would prefer to have, so that the information given is not overwhelming/useless.**”*

Lastly, the explainability graphs may have caused confusion among non-technical users as expressed by a participant (S71) who provided the following feedback about the experimental groups content-based recommender system: *“The straightforward recommendations were great, however **the data and graphs seem confusing for someone with limited knowledge of the subject.**”*

Following this, it is likely that the context-aware system lead to higher post-intervention motivation due to its superior usability and comprehensiveness. Table 40 contains samples of direct feedback from participants about their experiences with the context-aware recommender systems. The provided examples demonstrate that the experimental system may have had higher levels of exercise motivation due to the persuasive design strategies that focused on and emphasized user’s ability (self-efficacy). This was not emphasized in the control group.

Table 40: Subset of Open-ended feedback on the Context-Aware Recommender System (Separated by Study Group)

ID	Group	System	Comment	Feedback
S12	Control	Context-Aware	<i>“It recommended a lot of good exercises that I do incorporate into my gym sessions. The only issue I found was that it did give a lot of generic answers while I played around with it.”</i>	<i>“It would help if some answers weren’t so generic. I believe a better tailored answer would have been more impactful.”</i>
S16	Control	Context-Aware	<i>“Well, I liked the section wise answer. It recommended about exercises and diet separately. It uses very easy language and concise suggestions. That’s what I like about it. I would like if it could be a bit more specific actually and give me some more aspects.”</i>	<i>“I think it would be nice to get some more effective data from the users to better explain the situation. I think there is a field actually where you can put any kind of problems like food allergies or something like that, but still I think you can break it down a little bit or do something separately. also, like previously diagnosed problems and current medications.”</i>
S61	Control	Context-Aware	<i>“It was very simple to use. It asked for very little data and gave a plethora of results. However, sometimes it felt like it gave me the same recommendations regardless of my weight. Even providing a low weight it asked me to restrict my calories, when it should be asking for a calorie surplus. I do like that you can restrict the location of exercises (indoors vs outdoors) since many people have different circumstances.”</i>	<i>“I wished it tailored your routine based on your fitness goals, since mine is all about strength and size, whereas others might want to be losing fat.”</i>

Table 40 (continued)

ID	Group	System	Comment	Feedback
S38	Experimental	Context-Aware	<i>“I liked how it gave me other tips like how to stay motivated. The suggested food was not elaborate cooking.”</i>	<i>“An option to adjust goals like target calories a day and then get suggestions on the idea and a tailored meal would be great.”</i>
S65	Experimental	Context-Aware	<i>“The thing that stood out to me is that this system would be very beneficial as we will get every diet and exercise related info in 1 website.”</i>	<i>“Okay, so first when I put my information, it gave me good outputs. When I played around with it, the system gave me 1 motivational line that every step counts and that’s a thing that will push me so I would say whenever some one ask this system for recommendations, give some motivation to that person too.”</i>

The differences in self-efficacy and self-regulation between males and females is to be expected as gender differences are observed in the magnitude and types of intrinsic and extrinsic exercise motivation [105, 111]. Additionally, the positive correlation between self-efficacy and self-regulation is consistent with previous work which suggests how self-regulatory behaviours influence adherence to exercise plans by affecting self-efficacy [112, 113, 114, 115]. For example, a previous study suggests that teaching people how to change their behaviour first will increase their self-efficacy over time which also positively affects adherence to the new behaviour [113].

Lastly, it can be explained that the context-aware recommender systems offered the highest levels of projected post-intervention behaviour performance because of the positive correlation identified with post-intervention motivation through a Spearman’s Rho test ($\rho = 0.353$, $p < 0.001$). Usability, which is correlated with higher levels of exercise goals, adoption, and adherence [116, 117] may have positively impacted projected behaviour performance as well. For example, higher levels of system complexity may negatively affect technological adoption rates resulting in lower exercise scores for the content-based system [117].

Males in the control group reported significantly higher projected behaviour performance scores compared to the experimental group. The difference may be partly explained by the higher levels of post-intervention self-regulation afforded by the control systems (See Figure 32). Females in the experimental group also reported higher exercise scores through increased self-efficacy compared to males, indicating that persuasive design may have been more effective based on the type of gender. These findings are consistent with previous work which explores the gender-based effectiveness of persuasive design for health-based interventions [118].

While the results may not have indicated a significant group effect between recommender systems that leverage theory-driven frameworks from ad hoc design, the small but positive effect sizes reported from this study are consistent with other studies [28, 119]. The findings with privacy being a significant and positive construct for user experience was consistent with prior work that also leveraged the HAI-TIME model [74]. Altogether, the results point to promising future work which could investigate other psychosocial frameworks and techniques to make fitness recommender systems more personalized and tailored to a users' end goals and priorities.

6 Limitations and Future Work

In this section, the limitations of the study and system design are discussed and summarized from the researchers' and participants' perspectives.

6.1 Limitations in the Study Design

Firstly, the user study leveraged purposive (snowball [96]) and convenience (self-selection [97]) sampling to recruit as many participants as possible within a short period of time (May 12, 2025 - June 16, 2025). A post-hoc power analysis revealed that the findings of this Thesis were only capable of achieving a 55.8% probability for which ART ANOVA tests may correctly detect a significant effect based on the gender distributions with $\alpha, \beta = 0.05$. Altogether, the findings from the study design should be interpreted with caution.

After evaluating the participant demographics (See Chapter 4.1.1), it became clear that self-selection bias [120] may have impacted the study findings and generalizability. The percentage of participants who self-reported engaging in the recommended guidelines for weekly physical activity was 72.6% (53/73). This was much higher than what has been seen in related work for similar demographics; mainly, college students (40 – 50%) [116, 119]. The biasing of physically active participants may have also been caused by self-reporting bias to appear more socially desirable [121], however, the findings may still provide rich insights into the behavioural patterns of physically active participants. Future work can aim to address self-selection bias by using internal and external validation such as laboratory measures [121]. Additionally, future studies could broaden the eligibility requirements to include those with physical disabilities which represents an understudied study population [119].

Secondly, due to time constraints, the study design was selected to be cross-sectional over longitudinal. Perceptual measures on the dependent variables had to be selected since participants would not have access to the fitness-based recommender systems outside of the laboratory setting. Future work could consider the use of longitudinal studies as they will have stronger causal inferences, validity, and confidence intervals, while also reducing common method variance [122].

Thirdly, the effects of age, ethnicity, and level of education on the independent and dependent variables could not be studied due to small group sizes as seen in Appendix E.1. In general, a post-hoc power analysis based on the splits between gender and system type identified a low probability ($\leq 50\%$) of correctly detecting a significant effect for $\alpha = 0.05$. Due to the small groups, there was no meaningful way to collapse subgroups to obtain distinguishable results that generalize appropriately to other study populations of interest.

Fourthly, the study design only accounted for a small subset of psychosocial theories and frameworks which could impact exercise motivation and positive projected behaviour performance based on particular demographics such as gender and socio-economic status. Intrinsic and extrinsic motivation should be explicitly considered in future work as there are validated differences among males and females [116]. Dynamic frameworks and tools such as the Stages of Change Model [123], the Health Belief Model [124], or Motivational Interviewing [124] should be studied and integrated into future work, extending the study findings by considering different stages of exercise adoption and preparedness through validated cognitive

and behavioural processes (e.g., Consciousness Raising, Counterconditioning) [119, 124].

Lastly, the ordering of questions regarding current exercise behaviour without the recommender system and projected post-intervention behaviour performance with the recommender system may have introduced confounding effects during data collection in the questionnaire.

In spite of the limitations resulting from the study design, the results provide insights into how persuasive design affects the attitudes of active young adults in further improving their physical health. Given the small power size and small but positive outcomes, the results of this work are consistent in reliability and power with existing work. A larger number of participants is required to identify results that can be confidently stated.

6.2 Limitations in System Design

Firstly, the system design only supported two types of recommender systems; content-based and context-aware. While we did prepare a collaborative-filtering recommender system for evaluation, it could not be tested due to difficulty recruiting a sufficient number of participants. Future work should investigate the effectiveness of hybrid recommender systems as they are becoming more ubiquitous across various domains [74]. The use of multiple features that are both content-based, demographic, and context-aware could allow for better personalized and persuasive fitness plans which positively impact projected behaviour performance.

Secondly, the machine learning models used to support the content-based recommender systems were limited in complexity due to the availability of large and robust fitness datasets. A comprehensive search through open-access repositories such as Kaggle and Mendeley revealed little to no fitness-based recommendation datasets with the necessary features, quality, or quantity. Even after a relevant dataset was found and underwent filtering, imputation, scaling, and random oversampling to reduce noise [125], the model still showed signs of overfitting. Nevertheless, the results may still be of interest, especially for developing countries to design or implement fitness interventions where compute resources are heavily restricted by the digital divide [126].

Thirdly, the selection of a large language model used to support the personalized generation of fitness plans was restricted by access to sufficiently available hardware. Due to low video random access memory (VRAM), state-of-the-art models with reasoning capabilities such as Nemotron-4 340B [127] could not be used in the system pipelines. Nevertheless, the results may still be useful in countries where access to high-end compute is infeasible. Future work could look into testing different sizes of large language models on fitness recommendation tasks while also establishing empirical benchmarks on validated open-source datasets.

Lastly, the system only evaluated the use of a Zero Shot React Description Agent in the context-aware recommender system under deterministic parameters (e.g., Temperature). Testing different temperature values could yield better results that balance creativity with consistent high-quality fitness plans. Future work can investigate the effectiveness of other agents such as Conversational React, which have already been used in healthcare settings [128] to support multi-turn chat dialogues. Additionally, future work can aim to quantitatively assess the effectiveness of implicit versus explicit personalization on fitness-based recommender systems [128].

6.3 Limitations Reported by Participants

At the end of the study, participants were asked to provide open-ended feedback about how the recommender system could be improved in future iterations. The first author classified the sentiment classifications independently as the length and structure of the feedback deviated significantly from the dataset which was used to train Roberta Base Sentiment [100]. A majority of the comments were neutral (95.9%, 70/73), followed by negative (2.8%, 2/72) and positive (1.4%, 1/72). A Chi-Square test was undertaken to evaluate the relationship between the gender, group, and system variables with respect to sentiment counts. No significant effects ($p < 0.05$) were observed. Tables 41 and 42 summarize the key findings by theme and suggestions for improving the content-based and context-aware recommender systems.

Table 41: Identified Themes from Open-Ended Participant Feedback for Future Improvements to the Content-Based Recommender System

Theme	Improvements	ID	Comment
Usability	Allow users to select specific muscle groups and gym equipment by leveraging pictures or diagrams.	S63	<i>“I think that it could maybe include a photo with labels of all the muscle groups for people that might not know the names of all the muscle groups as well as where they are located [...]”</i>
Feature Augmentation	Allow users to generate a fitness plan based on factors such as availability during the week, dietary preferences, current fitness level, and existing health issues and medications.	S39	<i>“I would have it first gather data about the individuals starting point in their fitness journey and gather information about their bodies to know what they are and are not capable to perform when starting. I also think there should be a way to track progress and have the recommender adjust and update”</i>
Safety Alignment	Provide explicit step-by-step instructions on how to perform the recommended exercises with proper form. Add general safety recommendations to maximize exercise gains.	S69	<i>“[...] safety recommendations would be valuable so that users can know how to perform the exercise safely and properly with good form in order to reap the most benefit.”</i>

Table 41 (continued)

Theme	Improvements	ID	Comment
Functionality	Allow users to select multiple muscle groups and gym equipment instead of one at a time to reduce the number of runs on the application.	S56	<i>“I think that it would be best to be able to input multiple options as opposed to just one for each preference. If i were assembling an ”Upper Body Day”, I don’t have that ability, rather just focusing on individual muscle groups. Or if I have specific equipment or machines at my local gym, I can input those in a list, and have the recommender use that data to output possible exercises that I can perform with my given list of equipment.”</i>
Data Augmentation	Augment the dataset used to train the recommender system. Include sports, daily chores, and stretches into the set of recommendations (labels).	S11	<i>“[...] give examples of sports or daily tasks that would benefit from that muscle group being exercised.”</i>

Table 42: Identified Themes from Open-Ended Participant Feedback for Future Improvements to the Context-Aware Recommender System

Theme	Improvements	ID	Comment
Feature Augmentation	Allow users to select their goal and input existing habits such as calorie intake per day, availability, mood, and culture.	S55	<i>“I would suggest adding more questionnaires beyond age and height, such as hours spent at work/school, or lifestyle, routine, mental health etc. to cater to each individual’s specific needs when it comes to taking care of their physical health.”</i>

Table 42 (continued)

Theme	Improvements	ID	Comment
Broad Recommendations	Final fitness plans should include concrete numbers such as the number of sets and reps to perform. The explanations should address muscle activation as well.	S04	<i>“I think that the suggestions are great for what they accomplish with sedentary behaviours as the recommendations are very small easy changes that people can make in their lives. I think that having some more concrete numbers would be helpful as people can see what numbers they should aim for. However, I understand that this may be triggering to some people so it is a fine line.”</i>
Functionality	Add conversational and multi-modal support; include pictures or videos of active and sedentary lifestyles. Allow the user to continually adapt the plan. Use tables rather than relying on text-based paragraph suggestions. Lastly, empirical claims about weight loss or gain should be substantiated with citations.	S66	<i>“Rather than writing a long paragraph, maybe create a weekly time-table. Allow users to ask questions on the given response for more clarification. Allow user to write their current schedule so the new schedule could be build upon it with slow progression rather than providing a totally unfamiliar workout routine.”</i>

The findings are largely consistent with results obtained in Chapter 4.1.4. Irrespective of the group and type of recommender system, there is a need for feature augmentation and multimodal support to enhance user experience and the personalization of fitness plans. The sparsity of high-quality datasets with relevant features is consistent with the cold-start problem; a well-documented limitation associated with recommender systems in various domains [74]. Future work could investigate the effectiveness of applying feature and data augmentation on existing datasets by using large language models [127], social signals [53], or healthcare experts [54].

6.4 Implications

Upon analyzing the open-ended participant responses about self-regulation, their reflections were complex, relying on certain techniques with fundamental gaps in domain knowledge. A participant (S51) wrote, “*I am very goal orientated but I usually set more long term goals and don’t really know how to effectively break them up into weekly goals. Additionally I am not the type of person to share my progress unless I have reached a significant milestone.*” In trying to increase exercise self-regulation, the open-ended feedback demonstrated that being able to break a goal into discrete step-by-step actions is integral. Having a schedule without the knowledge of how to properly integrate it won’t result in increased physical activity, even if one has a high level of motivation going into their fitness journey [119]. The findings and themes are relatively consistent with previous work which identifies self-regulation as being mostly intrinsic and interpersonal in influencing long-term exercise motivation, self-efficacy, and adoption [129, 130, 113, 114, 115]. With these results, designers and evaluators of fitness-based recommender systems can start developing interventions which place more support on executive functioning skills that build and foster long-term motivation rather than only relying on self-efficacy to drive projected behaviour performance.

Additionally, there are implications for future user interface design. The findings demonstrate that text-based reasoning organized in headers and bullet point lists may not be optimal in effectively conveying the information. Open-ended feedback demonstrated the desire from participants (S43) to control visible headers and also organize fitness and dietary plans into tables. This would help the user quickly find key areas of interest while filtering out the unnecessary suggestions. There is a need to explore the effectiveness of communicating fitness informations to users based on their domain expertise and exercise preparedness.

Lastly, multimodal support through diagrams, pictures, and videos to identify and safely operate muscle groups and exercise equipment could be provided through the application of multi-agent systems [131] to support improved user experience and safety-alignment in fitness-based recommender systems. The recommender systems that were used in this work only used a single React Agent which may have lead to the associated feedback provided by participants. While the original prompts used to support the LangChain Agent could be improved, LangGraph [132] should still be explored in the implementation of a revised recommender system. Figure 35 provides an example workflow to address the limitations addressed by participants resulting from broad explanations with a desire for proper safety alignment and adequate features. The workflow could also be adapted to support a conversational agent which addresses the functionality requirements in Tables 41 and 42.

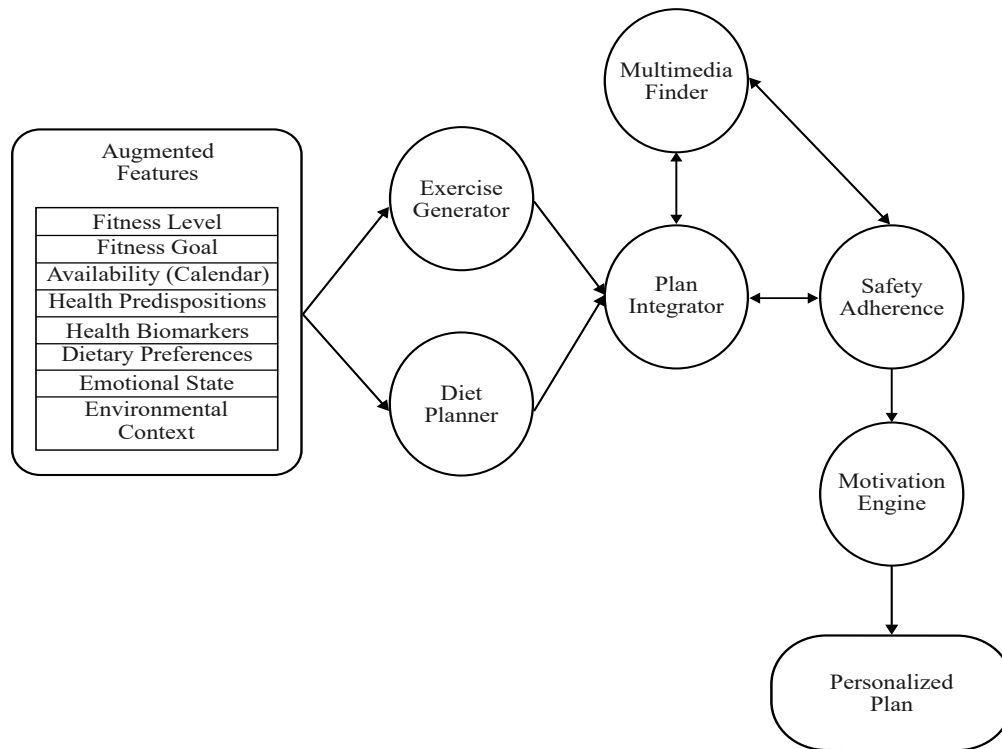


Figure 35: Proposed High-Level Multi-Agent System for Generating Personalized Fitness Plans. Nodes Represent Individual Agents and Edges Represent Communication Between Different Actors/Critics.

Investigating tools and frameworks which keep compute resources and costs low can assist in the development of sustainable and accessible health campaigns which aim to teach people about the harms and preventative measures of sedentary behaviour and obesity. Prior work has been undertaken to investigate this research, however, the certainty of evidence remains low to moderate [72, 59].

7 Conclusion

This thesis investigates how explainability, theory-driven psychosocial frameworks, and user experience can be leveraged into the design and evaluation of fitness-based recommender systems. A comprehensive literature review was undertaken (Chapter 2) and user feedback (Chapters 6.3) was analyzed to provide an overview of current trends, limitations, and areas for future work that researchers and practitioners can use to further personalize fitness interventions.

The results from our experimental user study demonstrated that participants characterized positive user experience with fitness-based recommender systems being more usable and privacy-preserving in contrast with other constructs from the HUED-CUP framework. The context-aware recommender systems, regardless of group, performed the best on all measures of the dependent variables. Higher perceived usability may have been associated with the context-aware condition due to its ability to also provide a diet plan, whereas the content-based recommender system only produced exercise plans.

The experimental context-aware recommender system had the highest post-intervention exercise motivation scores. Females in the experimental condition showed higher levels of self-efficacy compared to the control, however, the differences were numerical and not statistically-significant. The reverse held true for males. The control group reported higher levels of post-intervention projected behaviour performance compared to the experimental group. Females in the experimental group reported significantly higher projected behaviour performance scores compared to males in the same group. Males in the control group also reported significant higher projected behaviour performance scores compared to males in the experimental group. These results may be due to the high number of already physically active participants who volunteered to take part in the study which possibly skewed the findings under certain conditions.

Altogether, the results demonstrate that explainability, psychosocial frameworks, and user experience factors that were used to design the fitness-based recommender systems are perceived differently among males and females. Females were shown to benefit from the application of explainability and psychosocial frameworks whereas it made no significant difference for males in terms of motivation or projected behaviour performance.

User feedback identified that a personalized fitness-based recommender system must be able to consider injuries, weather, emotional state, and availability into the curation of a weekly plan. Furthermore, the plan must provide explicit instructions for routine integration and exercise completion. Variety of exercises and difficulty must be well-aligned based on levels of self-efficacy and fitness levels / goals. Lastly, the persons health conditions such as menstrual cycle for females needs to be considered.

This work contributes to the field by presenting results from an experimental user study which assesses the effectiveness of context-aware and content-based recommender systems that implement varying levels of explainability and projected behaviour performance techniques. Furthermore, gender-based differences are reported and participant feedback is obtained and summarized for future work. In particular, this work investigates how personal, cognitive, social, and environmental factors affect fitness motivation and behaviour, which has shown to be understudied from existing literature [119, 28]. Addressing exercise motivation is necessary to positively affecting projected behaviour performance as shown by the reported correlations from this study.

Bibliography

- [1] K. Oyibo, "Comtech: Towards a unified taxonomy of persuasive techniques for persuasive technology design," *Computers in Human Behavior Reports*, vol. 14, p. 100372, 2024.
- [2] J. C. Greene, V. J. Caracelli, and W. F. Graham, "Toward a conceptual framework for mixed-method evaluation designs," *Educational evaluation and policy analysis*, vol. 11, no. 3, pp. 255–274, 1989.
- [3] R. Murarka, "Gym exercises dataset."
- [4] A. F. Petrella, D. P. Gill, and R. J. Petrella, "Evaluation of the get active questionnaire in community-dwelling older adults," *Applied Physiology, Nutrition, and Metabolism*, vol. 43, no. 6, pp. 587–594, 2018.
- [5] A. E. Damin, R. Nitrini, and S. M. D. Brucki, "Cognitive change questionnaire as a method for cognitive impairment screening," *Dementia & Neuropsychologia*, vol. 9, no. 3, pp. 237–244, 2015.
- [6] S. P. Mullen, E. A. Olson, S. M. Phillips, A. N. Szabo, T. R. Wójcicki, E. L. Mailey, N. P. Gothe, J. T. Fanning, A. F. Kramer, and E. McAuley, "Measuring enjoyment of physical activity in older adults: invariance of the physical activity enjoyment scale (paces) across groups and time," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 8, pp. 1–9, 2011.
- [7] M. H. Chignell, A. Quan-Haase, and J. Gwizdka, "The privacy attitudes questionnaire (paq): initial development and validation," in *proceedings of the human factors and ergonomics society annual meeting*, vol. 47, pp. 1326–1330, SAGE Publications Sage CA: Los Angeles, CA, 2003.
- [8] B. Resnick and L. S. Jenkins, "Testing the reliability and validity of the self-efficacy for exercise scale," *Nursing research*, vol. 49, no. 3, pp. 154–159, 2000.
- [9] L. S. Rovniak, E. S. Anderson, R. A. Winett, and R. S. Stephens, "Social cognitive determinants of physical activity in young adults: a prospective structural equation analysis," *Annals of behavioral medicine*, vol. 24, pp. 149–156, 2002.
- [10] B. J. Fogg, "A behavior model for persuasive design," in *Proceedings of the 4th international Conference on Persuasive Technology*, pp. 1–7, 2009.
- [11] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive systems design: Key issues, process model, and system features," *Communications of the association for Information Systems*, vol. 24, no. 1, p. 28, 2009.
- [12] A. Bandura, "Social cognitive theory of self-regulation," *Organizational behavior and human decision processes*, vol. 50, no. 2, pp. 248–287, 1991.
- [13] A. Luszczynska and R. Schwarzer, "Social cognitive theory," *Fac Health Sci Publ*, vol. 2015, pp. 225–251, 2015.

- [14] S. Michie, M. M. Van Stralen, and R. West, “The behaviour change wheel: a new method for characterising and designing behaviour change interventions,” *Implementation science*, vol. 6, pp. 1–12, 2011.
- [15] T. J. Willmott, B. Pang, and S. Rundle-Thiele, “Capability, opportunity, and motivation: an across contexts empirical examination of the com-b model,” *BMC Public Health*, vol. 21, no. 1, p. 1014, 2021.
- [16] S. S. Sundar, “Rise of machine agency: A framework for studying the psychology of human–ai interaction (haii),” *Journal of Computer-Mediated Communication*, vol. 25, no. 1, pp. 74–88, 2020.
- [17] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” *Journal of management information systems*, vol. 24, no. 3, pp. 45–77, 2007.
- [18] A. Bandura, “Self-efficacy: Toward a unifying theory of behavioral change,” *Advances in behaviour research and therapy*, vol. 1, no. 4, pp. 139–161, 1978.
- [19] K. Oyibo, *Designing culture-tailored persuasive technology to promote physical activity*. PhD thesis, University of Saskatchewan, 2020.
- [20] D. Keyson and K. Parsons, “Designing the user interface using rapid prototyping,” *Applied ergonomics*, vol. 21, no. 3, pp. 207–213, 1990.
- [21] R. C. Williges, “The society’s lecture 1987 the use of models in human-computer interface design,” *Ergonomics*, vol. 30, no. 3, pp. 491–502, 1987.
- [22] R. Burke, A. Felfernig, and M. H. Göker, “Recommender systems: An overview,” *Ai Magazine*, vol. 32, no. 3, pp. 13–18, 2011.
- [23] B. Smyth, “Recommender systems: A healthy obsession,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 9790–9794, 2019.
- [24] M. Wiesner and D. Pfeifer, “Health recommender systems: concepts, requirements, technical basics and challenges,” *International journal of environmental research and public health*, vol. 11, no. 3, pp. 2580–2607, 2014.
- [25] X. Liu, B. Gao, B. Suleiman, H. You, Z. Ma, Y. Liu, and A. Anaissi, “Privacy-preserving personalized fitness recommender system p3fitrec: a multi-level deep learning approach,” *ACM Transactions on Knowledge Discovery from Data*, vol. 17, no. 6, pp. 1–24, 2023.
- [26] I. Coppens, L. Martens, and T. De Pessemier, “Analyzing accuracy versus diversity in a health recommender system for physical activities: a longitudinal user study,” in *Proceedings of the 17th ACM Conference on Recommender Systems*, pp. 1146–1151, 2023.

- [27] J. P. Agans, F. Ma, S. Schade, and C. Sciamanna, “Supporting physical activity adoption through recommender system technology: A pilot study,” *Journal of Health Psychology*, p. 13591053241242541, 2024.
- [28] S. Toyonaga and K. Oyibo, “An analysis of recommender systems within the fitness and exercise domain: Systematic review.” Unpublished Manuscript, 2024.
- [29] P. Venkatachalam and S. Ray, “How do context-aware artificial intelligence algorithms used in fitness recommender systems? a literature review and research agenda,” *International Journal of Information Management Data Insights*, vol. 2, no. 2, p. 100139, 2022.
- [30] R. Ali, M. Afzal, M. Sadiq, M. Hussain, T. Ali, S. Lee, and A. M. Khattak, “Knowledge-based reasoning and recommendation framework for intelligent decision making,” *Expert Systems*, vol. 35, no. 2, p. e12242, 2018.
- [31] E. Martinez-Martin, A. Costa, and M. Cazorla, “Pharos 2.0—a physical assistant robot system improved,” *Sensors*, vol. 19, no. 20, p. 4531, 2019.
- [32] P. J. Giabbanelli and R. Crutzen, “Supporting self-management of obesity using a novel game architecture,” *Health informatics journal*, vol. 21, no. 3, pp. 223–236, 2015.
- [33] K. Oyibo and J. Vassileva, “Gender preference and difference in behavior modeling in fitness applications: A mixed-method approach,” *Multimodal Technologies and Interaction*, vol. 4, no. 2, p. 21, 2020.
- [34] T. Lehnert, D. Sonntag, A. Konnopka, S. Riedel-Heller, and H.-H. König, “Economic costs of overweight and obesity,” *Best practice & research Clinical endocrinology & metabolism*, vol. 27, no. 2, pp. 105–115, 2013.
- [35] A. del Rio, J. Jimenez, R. Medina-Garcia, C. Lozano-Hernandez, F. Alvarez, and J. Serrano, “Improving quality of life in chronic patients: A pilot study on the effectiveness of a health recommender system and its usability,” *Applied Sciences*, vol. 13, no. 10, p. 5850, 2023.
- [36] P. T. James, R. Leach, E. Kalamara, and M. Shayeghi, “The worldwide obesity epidemic,” *Obesity research*, vol. 9, no. S11, pp. 228S–233S, 2001.
- [37] J. M. Rippe, S. Crossley, and R. Ringer, “Obesity as a chronic disease: modern medical and lifestyle management,” *Journal of the American Dietetic Association*, vol. 98, no. 10, pp. S9–S15, 1998.
- [38] I. Mosby and T. Galloway, ““hunger was never absent”: How residential school diets shaped current patterns of diabetes among indigenous peoples in Canada,” *Cmaj*, vol. 189, no. 32, pp. E1043–E1045, 2017.
- [39] P. B. Sparling, “Obesity on campus,” *Preventing chronic disease*, vol. 4, no. 3, 2007.

- [40] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, “The aligned rank transform for nonparametric factorial analyses using only anova procedures,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 143–146, 2011.
- [41] L. A. Elkin, M. Kay, J. J. Higgins, and J. O. Wobbrock, “An aligned rank transform procedure for multifactor contrast tests,” in *The 34th annual ACM symposium on user interface software and technology*, pp. 754–768, 2021.
- [42] T. Toscos, S. Consolvo, and D. W. McDonald, “Barriers to physical activity: a study of self-revelation in an online community,” *Journal of medical systems*, vol. 35, pp. 1225–1242, 2011.
- [43] A. Jaiswal, G. Chauhan, and N. Srivastava, “Using learnable physics for real-time exercise form recommendations,” in *Proceedings of the 17th ACM Conference on Recommender Systems*, pp. 688–695, 2023.
- [44] A. Mahyari and P. Pirolli, “Physical exercise recommendation and success prediction using interconnected recurrent neural networks,” in *2021 IEEE International Conference on Digital Health (ICDH)*, pp. 148–153, IEEE, 2021.
- [45] I. Kulev, C. Valk, Y. Lu, and P. Pu, “Recommender system for responsive engagement of senior adults in daily activities,” *Journal of Population Ageing*, vol. 13, pp. 167–185, 2020.
- [46] I. Coppens, L. Martens, and T. De Pessemer, “Motivating people to move more with personalized activity and tip recommendations: A randomized controlled trial,” in *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, pp. 123–126, 2023.
- [47] I. Coppens, T. De Pessemer, and L. Martens, “Connecting physical activity with context and motivation: a user study to define variables to integrate into mobile health recommenders,” *User Modeling and User-Adapted Interaction*, vol. 34, no. 1, pp. 147–181, 2024.
- [48] E. Z. Tragos, D. O’Reilly-Morgan, J. Geraci, B. Shi, B. Smyth, C. Doherty, A. Lawlor, and N. Hurley, “Keeping people active and healthy at home using a reinforcement learning-based fitness recommendation framework,” in *IJCAI*, pp. 6237–6245, 2023.
- [49] S. Ono, Y. Yotsuya, N. Takahashi, T. Sakamoto, and T. Kato, “Physical-exercise recommendation system based on individual daily schedule,” in *2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS)*, pp. 1–3, IEEE, 2022.
- [50] Y. Lin, J. Jessurun, B. De Vries, and H. Timmermans, “Motivate: Towards context-aware recommendation mobile system for healthy living,” in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pp. 250–253, IEEE, 2011.
- [51] Z. Li, S. Das, J. Codella, T. Hao, K. Lin, C. Maduri, and C.-H. Chen, “An adaptive, data-driven personalized advisor for increasing physical activity,” *IEEE journal of biomedical and health informatics*, vol. 23, no. 3, pp. 999–1010, 2018.

- [52] S. Sengan, V. Subramaniaswamy, R. H. Jhaveri, V. Varadarajan, R. Setiawan, and L. Ravi, “A secure recommendation system for providing context-aware physical activity classification for users,” *Security and Communication Networks*, vol. 2021, no. 1, p. 4136909, 2021.
- [53] S. Dharia, M. Eirinaki, V. Jain, J. Patel, I. Varlamis, J. Vora, and R. Yamauchi, “Social recommendations for personalized fitness assistance,” *Personal and Ubiquitous Computing*, vol. 22, pp. 245–257, 2018.
- [54] A. Mahyari, P. Pirolli, and J. A. LeBlanc, “Real-time learning from an expert in deep recommendation systems with application to mhealth for physical exercises,” *IEEE journal of biomedical and health informatics*, vol. 26, no. 8, pp. 4281–4290, 2022.
- [55] H. Alcaraz-Herrera and I. Palomares, “Evolutionary approach for ‘healthy bundle’ wellbeing recommendations.,” in *HealthRecSys@ RecSys*, pp. 18–23, 2019.
- [56] D. Chowdhury, A. Roy, S. R. Ramamurthy, and N. Roy, “Charlie: A chatbot that recommends daily fitness and diet plans,” in *2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, pp. 116–121, IEEE, 2023.
- [57] P. Pilloni, L. Piras, L. Boratto, S. Carta, G. Fenu, F. Mulas, *et al.*, “Recommendation in persuasive ehealth systems: An effective strategy to spot users’ losing motivation to exercise,” in *CEUR Workshop Proceedings*, vol. 1953, pp. 6–9, 2017.
- [58] N. Kadri, A. Ellouze, and M. Ksantini, “Recommendation system for human physical activities using smartphones,” in *2020 2nd International Conference on Computer and Information Sciences (ICCIS)*, pp. 1–4, IEEE, 2020.
- [59] Y. Takama, W. Sasaki, T. Okumura, C.-C. Yu, L.-H. Chen, and H. Ishikawa, “Walking route recommender for supporting a walk as health promotion,” *IEICE TRANSACTIONS on Information and Systems*, vol. 100, no. 4, pp. 671–681, 2017.
- [60] M. MINAKATA, T. MARUYAMA, M. TADA, and Y. KURITA, “Safe walking route recommender based on fall risk calculation using digital human model on 3d map,” *IEEE Access*, vol. 10, pp. 1–6, 2021.
- [61] P. Palumbo, L. Cattelani, F. Fusco, M. Pijnappels, L. Chiari, F. Chesani, and S. Mellone, “A recommender system for behavioral change in 60-70-year-old adults,” in *CEUR Workshop Proceedings*, vol. 2804, pp. 69–80, 2020.
- [62] S. Wongpun and S. Guha, “Elderly care recommendation system for informal caregivers using case-based reasoning,” in *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 548–552, IEEE, 2017.
- [63] T. T. Tran, J. W. Choi, C. Van Dang, G. SuPark, J. Y. Baek, and J. W. Kim, “Recommender system with artificial intelligence for fitness assistance system,” in *2018 15th International Conference on Ubiquitous Robots (UR)*, pp. 489–492, IEEE, 2018.

- [64] Z. Zhao, A. Arya, R. Orji, and G. Chan, "Physical activity recommendation for exergame player modeling using machine learning approach," in *2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH)*, pp. 1–9, IEEE, 2020.
- [65] J. Ichino, M. Miyazaki, and K. Tanaka, "How personalization algorithm changes consumer behaviors in the digital healthcare service," in *Leveraging Transdisciplinary Engineering in a Changing and Connected World*, pp. 3–12, IOS Press, 2023.
- [66] Z. Zhao, A. Arya, R. Orji, G. Chan, *et al.*, "Effects of a personalized fitness recommender system using gamification and continuous player modeling: system design and long-term validation study," *JMIR serious games*, vol. 8, no. 4, p. e19968, 2020.
- [67] M. Afzal, S. I. Ali, R. Ali, M. Hussain, T. Ali, W. A. Khan, M. B. Amin, B. H. Kang, and S. Lee, "Personalization of wellness recommendations using contextual interpretation," *Expert Systems with Applications*, vol. 96, pp. 506–521, 2018.
- [68] Y. Yang and Y. Zhao, "Personalized sports health recommendation system assisted by q-learning algorithm," *International Journal of Human–Computer Interaction*, pp. 1–13, 2024.
- [69] R. Ali, M. Afzal, M. Hussain, M. Ali, M. H. Siddiqi, S. Lee, and B. H. Kang, "Multimodal hybrid reasoning methodology for personalized wellbeing services," *Computers in biology and medicine*, vol. 69, pp. 10–28, 2016.
- [70] C. Basnayake, C. Peiris, H. Wickramaratna, and P. Jayathunga, "Recommender system based on food and exercise ontologies to find the suitable fitness exercise plan with the aid of python," in *2021 5th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI)*, pp. 1–6, IEEE, 2021.
- [71] I. Palomares, H. Alcaraz-Herrera, and K.-Y. Shen, "F-evorecsys: An extended framework for personalized well-being recommendations guided by fuzzy inference and evolutionary computing," *International Journal of Fuzzy Systems*, vol. 24, no. 6, pp. 2783–2797, 2022.
- [72] N. Gomez-del Rio, C. S. Gonzalez-Gonzalez, P. A. Toledo-Delgado, V. Munoz-Cruz, and F. Garcia-Penalvo, "Health promotion for childhood obesity: An approach based on self-tracking of data," *Sensors*, vol. 20, no. 13, p. 3778, 2020.
- [73] H. Alcaraz-Herrera, J. Cartlidge, Z. Toumpakari, M. Western, and I. Palomares, "Evorecsys: Evolutionary framework for health and well-being recommender systems," *User Modeling and User-Adapted Interaction*, vol. 32, no. 5, pp. 883–921, 2022.
- [74] Y. Sun, M. Drivas, M. Liao, and S. S. Sundar, "When recommender systems snoop into social media, users trust them less for health advice," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2023.
- [75] M. Balpande, J. Sharma, A. Nair, M. Khandelwal, and S. Dhanray, "Ai based gym trainer and diet recommendation system," in *2023 IEEE 4th Annual Flagship India Council International Subsections Conference (INDISCON)*, pp. 1–7, IEEE, 2023.

- [76] F. Gasparetti, L. M. Aiello, and D. Quercia, “Personalized weight loss strategies by mining activity tracker data,” *User Modeling and User-Adapted Interaction*, vol. 30, no. 3, pp. 447–476, 2020.
- [77] J. Berndsen, A. Lawlor, and B. Smyth, “Running with recommendation.,” in *HealthRecSys@ RecSys*, pp. 18–21, 2017.
- [78] J. Berndsen, B. Smyth, and A. Lawlor, “Fit to run: Personalised recommendations for marathon training,” in *Proceedings of the 14th ACM Conference on Recommender Systems*, pp. 480–485, 2020.
- [79] C. Feely, B. Caulfield, A. Lawlor, and B. Smyth, “Providing explainable race-time predictions and training plan recommendations to marathon runners,” in *Proceedings of the 14th ACM Conference on Recommender Systems*, pp. 539–544, 2020.
- [80] M. Donciu, M. Ionita, M. Dascalu, and S. Trausan-Matu, “The runner–recommender system of workout and nutrition for runners,” in *2011 13th international symposium on symbolic and numeric algorithms for scientific computing*, pp. 230–238, IEEE, 2011.
- [81] L. R. Ferretto, E. A. Bellei, D. Biduski, L. C. P. Bin, M. M. Moro, C. R. Cervi, and A. C. B. De Marchi, “A physical activity recommender system for patients with arterial hypertension,” *IEEE Access*, vol. 8, pp. 61656–61664, 2020.
- [82] F. R. Cavallo and C. Toumazou, “Personalised lifestyle recommendations for type 2 diabetes: Design and simulation of a recommender system on uk biobank data,” *PLOS Digital Health*, vol. 2, no. 8, p. e0000333, 2023.
- [83] V. S. Vairale and S. Shukla, “Physical fitness recommender framework for thyroid patients using restricted boltzmann machines.,” *International Journal of Intelligent Engineering & Systems*, vol. 13, no. 5, 2020.
- [84] S. Alian, J. Li, and V. Pandey, “A personalized recommendation system to support diabetes self-management for american indians,” *IEEE Access*, vol. 6, pp. 73041–73051, 2018.
- [85] J. Xie and Q. Wang, “A personalized diet and exercise recommender system for type 1 diabetes self-management: An in silico study,” *Smart Health*, vol. 13, p. 100069, 2019.
- [86] Y.-S. Chang, M. Han, B. Jeon, J.-C. Kim, and N. Park, “An neural collaborative filtering (ncf) based recommender system for personalized rehabilitation exercises,” in *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1292–1297, IEEE, 2023.
- [87] C. Gmez-Portes, J. J. Castro-Schez, J. Albusac, D. N. Monekosso, and D. Vallejo, “A fuzzy recommendation system for the automatic personalization of physical rehabilitation exercises in stroke patients,” *Mathematics*, vol. 9, no. 12, p. 1427, 2021.
- [88] C. S. Gonzalez-Gonzalez, P. A. Toledo-Delgado, V. Munoz-Cruz, and P. V. Torres-Carrion, “Serious games for rehabilitation: Gestural interaction in personalized gamified exercises through a recommender system,” *Journal of biomedical informatics*, vol. 97, p. 103266, 2019.

- [89] M. T. Ishraque, N. Zjalic, P. M. Zadeh, Z. Kobti, and P. Olla, "Artificial intelligence-based cardiac rehabilitation therapy exercise recommendation system," in *2018 IEEE MIT Undergraduate Research Technology Conference (URTC)*, pp. 1–5, IEEE, 2018.
- [90] J.-M. Blasco, M. Navarro-Bosch, J.-E. Aroca-Navarro, D. Hernández-Guillén, P. Puigcerver-Aranda, and S. Roig-Casasús, "A virtual assistant to guide early postoperative rehabilitation after reverse shoulder arthroplasty: A pilot randomized trial," *Bioengineering*, vol. 11, no. 2, p. 152, 2024.
- [91] O. Banos, M. B. Amin, W. A. Khan, T. Ali, M. Afzal, B. H. Kang, and S. Lee, "Mining minds: an innovative framework for personalized health and wellness support," in *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pp. 1–8, IEEE, 2015.
- [92] K. M. Kapp, "What is gamification," *The gamification of learning and instruction: game-based methods and strategies for training and education*, pp. 1–23, 2012.
- [93] D.-H. Shin and F. Biocca, "Health experience model of personal informatics: The case of a quantified self," *Computers in Human Behavior*, vol. 69, pp. 62–74, 2017.
- [94] K. Oyibo and S. Toyonaga, "Conceptual frameworks for designing and evaluating persuasive messages aimed at changing behavior: Systematic review," *Computers in Human Behavior Reports*, p. 100448, 2024.
- [95] J. Schoonenboom and R. B. Johnson, "How to construct a mixed methods research design," *Kolner Zeitschrift für Soziologie und Sozialpsychologie*, vol. 69, no. Suppl 2, p. 107, 2017.
- [96] K. Leighton, S. Kardong-Edgren, T. Schneidereith, and C. Foisy-Doll, "Using social media and snowball sampling as an alternative recruitment strategy for research," *Clinical simulation in nursing*, vol. 55, pp. 37–42, 2021.
- [97] L. Kai and N. R. Prabhala, "Self-selection models in corporate finance," *Handbook of empirical corporate finance*, pp. 37–86, 2007.
- [98] D. Kendzierski and K. J. DeCarlo, "Physical activity enjoyment scale: two validation studies.," *Journal of sport & exercise psychology*, vol. 13, no. 1, 1991.
- [99] E. McAuley, G. J. Jerome, D. X. Marquez, S. Elavsky, and B. Blissmer, "Exercise self-efficacy in older adults: social, affective, and behavioral influences," *Annals of behavioral medicine*, vol. 25, no. 1, pp. 1–7, 2003.
- [100] A. Joshy and S. Sundar, "Analyzing the performance of sentiment analysis using bert, distilbert, and roberta," in *2022 IEEE international power and renewable energy conference (IPRECON)*, pp. 1–6, IEEE, 2022.
- [101] L. Yu, R. Zhou, R. Chen, and K. K. Lai, "Missing data preprocessing in credit classification: One-hot encoding or imputation?," *Emerging Markets Finance and Trade*, vol. 58, no. 2, pp. 472–482, 2022.

- [102] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," in *2020 11th international conference on information and communication systems (ICICS)*, pp. 243–248, IEEE, 2020.
- [103] A. Kansal, "Chains, tools and agents," in *Building Generative AI-Powered Apps: A Hands-on Guide for Developers*, pp. 41–57, Springer, 2024.
- [104] Y. J. Yang, "An overview of current physical activity recommendations in primary care," *Korean journal of family medicine*, vol. 40, no. 3, p. 135, 2019.
- [105] E. S. Edwards and S. C. Sackett, "Psychosocial variables related to why women are less active than men and related health implications: supplementary issue: health disparities in women," *Clinical Medicine Insights: Women's Health*, vol. 9, pp. CMWH–S34668, 2016.
- [106] M. Mazziotta and A. Pareto, "Everything you always wanted to know about normalization (but were afraid to ask)," *Rivista Italiana di Economia Demografia e Statistica*, pp. 41–52, 2021.
- [107] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *biometrics*, pp. 159–174, 1977.
- [108] E. You, N. T. Lautenschlager, C. S. Wan, A. M. Goh, E. Curran, T. W. Chong, K. J. Anstey, F. Hanna, and K. A. Ellis, "Ethnic differences in barriers and enablers to physical activity among older adults," *Frontiers in Public Health*, vol. 9, p. 691851, 2021.
- [109] C. Thogersen-Ntoumani, M. Kritz, A. Grunseit, J. Chau, M. Ahmadi, A. Holtermann, A. Koster, C. Tudor-Locke, N. Johnson, C. Sherrington, *et al.*, "Barriers and enablers of vigorous intermittent lifestyle physical activity (vilpa) in physically inactive adults: a focus group study," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 20, no. 1, p. 78, 2023.
- [110] M. Dragoni, I. Donadello, and C. Eccher, "Explainable ai meets persuasiveness: Translating reasoning results into behavioral change advice," *Artificial Intelligence in Medicine*, vol. 105, p. 101840, 2020.
- [111] L. R. Duncan, C. R. Hall, P. M. Wilson, and O. Jenny, "Exercise motivation: a cross-sectional analysis examining its relationships with frequency, intensity, and duration of exercise," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 7, pp. 1–9, 2010.
- [112] E. McAuley, S. P. Mullen, A. N. Szabo, S. M. White, T. R. Wójcicki, E. L. Mailey, N. P. Gothe, E. A. Olson, M. Voss, K. Erickson, *et al.*, "Self-regulatory processes and exercise adherence in older adults," *Am J Prev Med*, vol. 41, no. 3, pp. 284–290, 2011.
- [113] J. J. Annesi and S. Gorjala, "Relations of self-regulation and self-efficacy for exercise and eating and bmi change: A field investigation," *BioPsychoSocial Medicine*, vol. 4, no. 1, p. 10, 2010.
- [114] J. Middelkamp, M. van Rooijen, P. Wolfhagen, and B. Steenbergen, "The effects of two self-regulation interventions to increase self-efficacy and group exercise behavior in fitness clubs," *Journal of sports science & medicine*, vol. 15, no. 2, p. 358, 2016.

- [115] N. Purdie and A. McCrindle, “Self-regulation, self-efficacy and health behavior change in older adults,” *Educational gerontology*, vol. 28, no. 5, pp. 379–400, 2002.
- [116] T. Egli, H. W. Bland, B. F. Melton, and D. R. Czech, “Influence of age, sex, and race on college students’ exercise motivation of physical activity,” *Journal of American college health*, vol. 59, no. 5, pp. 399–406, 2011.
- [117] L. Lun, D. Zetian, T. W. Hoe, X. Juan, D. Jiabin, and W. Fulai, “Factors influencing user intentions on interactive websites: Insights from the technology acceptance model,” *IEEE Access*, 2024.
- [118] R. Orji, “Exploring the persuasiveness of behavior change support strategies and possible gender differences,” *BCSS@ PERSUASIVE*, vol. 1153, pp. 41–57, 2014.
- [119] X. D. Keating, J. Guan, J. C. Piñero, and D. M. Bridges, “A meta-analysis of college students’ physical activity behaviors,” *Journal of American college health*, vol. 54, no. 2, pp. 116–126, 2005.
- [120] T. Hiratsuka, “The volunteer participation paradox: Ethical tensions between self-selection and targeted sampling,” *Research Methods in Applied Linguistics*, vol. 4, no. 2, p. 100206, 2025.
- [121] A. Althubaiti, “Information bias in health research: definition, pitfalls, and adjustment methods,” *Journal of multidisciplinary healthcare*, pp. 211–217, 2016.
- [122] A. Rindfleisch, A. J. Malter, S. Ganesan, and C. Moorman, “Cross-sectional versus longitudinal survey research: Concepts, findings, and guidelines,” *Journal of marketing research*, vol. 45, no. 3, pp. 261–279, 2008.
- [123] J. J. Hammermeister, R. M. Page, and D. Dolny, “Psychosocial, behavioral, and biometric characteristics of stages of exercise adoption,” *Psychological Reports*, vol. 87, no. 1, pp. 205–217, 2000.
- [124] B. C. Bock, S. E. Linke, M. A. Napolitano, M. M. Clark, R. B. Gaskins, B. H. Marcus, J. L. Van Raalte, and B. W. Brewer, “Exercise initiation, adoption, and maintenance in adults: Theoretical models and empirical support,” in *Exploring sport and exercise psychology*, pp. 165–189, Washington, DC: American Psychological Association, 3rd ed. ed., 2014.
- [125] K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, 2022.
- [126] E. M. Rogers, “The digital divide,” *Convergence*, vol. 7, no. 4, pp. 96–111, 2001.
- [127] B. Adler, N. Agarwal, A. Aithal, D. H. Anh, P. Bhattacharya, A. Brundyn, J. Casper, B. Catanzaro, S. Clay, J. Cohen, *et al.*, “Nemotron-4 340b technical report,” *arXiv preprint arXiv:2406.11704*, 2024.
- [128] A. B. Kocaballi, S. Berkovsky, J. C. Quiroz, L. Laranjo, H. L. Tong, D. Rezazadegan, A. Briatore, and E. Coiera, “The personalization of conversational agents in health care: systematic review,” *Journal of medical Internet research*, vol. 21, no. 11, p. e15360, 2019.

- [129] A. G. Box, Y. Feito, C. Brown, and S. J. Petruzzello, “Individual differences influence exercise behavior: how personality, motivation, and behavioral regulation vary among exercise mode preferences,” *Heliyon*, vol. 5, no. 4, 2019.
- [130] P. J. Teixeira, E. V. Carraca, D. Markland, M. N. Silva, and R. M. Ryan, “Exercise, physical activity, and self-determination theory: a systematic review,” *International journal of behavioral nutrition and physical activity*, vol. 9, pp. 1–30, 2012.
- [131] A. Dorri, S. S. Kanhere, and R. Jurdak, “Multi-agent systems: A survey,” *Ieee Access*, vol. 6, pp. 28573–28593, 2018.
- [132] V. Mavroudis, “Langchain v0. 3,” *Preprints*, 2024.

Appendices

A User Study (Instruments)

A.1 Consent Form

Evaluation of Fitness-Based Recommender System: Pre-Screening Questionnaire

Researchers

Kiemute Oyibo (PI), Lassonde School of Engineering, 4751 Keele Street (LRC 122),

kiemute.oyibo@yorku.ca

Shogo Toyonaga (MSc Student), Lassonde School of Engineering, 4751 Keele Street (LRC 122),

stoyonag@yorku.ca

Purpose of the Research

The objective of this study is to determine the perceptual effectiveness of fitness-based recommender systems. We seek to identify whether the use of theoretical frameworks grounded in social science and psychology in addition to explainability can help increase exercise adoption in young adults.

Please note that participation is not a requirement of the course and as such, it will not affect your grade or relationship with the course instructor and teaching staff in any way. Furthermore, students who are known to the researchers (Shogo Toyonaga, Kiemute Oyibo) through a prior relationship (e.g., student-teacher) will be ineligible to participate in the study to reduce subject bias.

What You Will Be Asked to Do in the Research

This is an in-person study and as such, you will be required to visit the Lassonde Research Center. Participants will be asked to start filling out a questionnaire prior to interacting with the fitness-based recommender system. This will measure baseline metrics about the participants motivation and privacy preferences. Following this, the participant will interact with the recommender system for 10-15 minutes, generating personalized fitness activities from each of the supported routes.

Given information about the participants' personal preferences and demographics, relevant exercises such as running, jogging, bench press, etc., will be provided. Participants will fill in the remaining part of the preliminary survey which gathers their thoughts about the effectiveness of the recommendations through user experience metrics (e.g., relevance, diversity). All participants who enroll in the study will receive \$20.00 (CAD).

You will not be asked to engage in any of the recommended fitness activities or plans. You should speak to a Health Physician first prior to engaging in new exercises to ensure your physical safety.

Risks and Discomforts

We do not foresee any risks or discomfort from your participation in the research.

Benefits of the Research and Benefits to You

Participants may benefit from the research by obtaining likeable and relevant exercise recommendations that can be incorporated into their daily schedules. Furthermore, the scholarly community may receive benefits from this research in the form of (1) facilitating University programs related to fostering healthy behaviours (e.g., exercise, diet) at a very low cost, (2) extending machine learning research from a sociotechnical perspective, and (3) determining whether classical machine learning and generative AI can overcome the one-size fits-all approach.

Voluntary Participation and Withdrawal

Your participation in the study is completely voluntary and you may choose to stop participating at any time. Your decision not to volunteer, to stop participating, or to refuse to answer particular questions will not influence (1) the treatment you may be receiving, (2) the nature of the ongoing relationship you may have with the researchers or study staff, and (3) the nature of your relationship with York University either now, or in the future.

If you stop participating, you will still be eligible to receive the \$20.00 (CAD) for agreeing to be in the project, even if you withdraw without completion of the research. In the event you withdraw from the study, all associated data collected will be immediately destroyed wherever possible.

Confidentiality

Unless you choose otherwise all the information you supply during the research will be held in confidence. Additionally, unless you specifically indicate your consent, your name will not appear in any report or publication of the research.

Your data will only be collected through questionnaires. Your data will be safely stored on a secure Google Drive and encrypted hard drive during the study. Only research staff/research team members will have access to this information. Confidentiality will be fully provided possible by law.

The data collected in this research project may be used – in an anonymized form - by members of the research team in subsequent research investigations exploring similar lines of inquiry. Such projects will still undergo ethics review by the HPRC, our institutional REB. Any secondary use of anonymized data by the research team will be treated with the same degree of confidentiality and anonymity as in the original research project.

Please note that at the end of the study, the anonymized data may be deposited into one or more publicly accessible scientific repositories, such as York University Dataverse, so that data may be inspected and analyzed by other researchers. The data that will be shared on York University Dataverse (Borealis) will not contain any information that can identify you. It will be held under a CC BY-NC 4.0 open license.

Please note that the survey(s) are being conducted with the help of Google Forms, a company not affiliated with York University and with its own privacy and security policies that you can find at

<https://policies.google.com/>. There is always a risk during web-based transmission that data can be intercepted by a third party (e.g., government agencies, hackers) and thus the confidentiality and privacy cannot be guaranteed.

Questions About the Research?

If you have questions about the research in general or about your role in the study, please feel free to contact Kiemute Oyibo by e-mail (kiemute.oyibo@yorku.ca). You may also reach out to the Electrical Engineering and Computer Science (EECS) Graduate program by telephone at (416)-650-8215, or by email (gpa@eecs.yorku.ca).

This research has received ethics review and approval (#e2025-062) by the Human Participants Review Sub-Committee, York University’s Ethics Review Board and conforms to the standards of the Canadian Tri-Council Research Ethics guidelines. If you have any questions about this process, or about your rights as a participant in the study, please contact the Manager, Research Ethics in the Office of Research Ethics, York University (e-mail ore@yorku.ca). This office oversees the ethical conduct of research studies and is not part of the study team. Everything that you discuss will be kept confidential.

A.2 Pre-screening Questionnaire

Evaluation of Fitness-Based Recommender System: Pre-Screening Questionnaire

1. Are you currently between the ages of 18-30 years old?
 - Yes
 - No

2. Have you experienced any of the following symptoms or diagnoses within the last year?

Table A.2.1: Get Active Questionnaire [4]

Prompt	Please select all that apply
A diagnosis or treatment for a condition (e.g., cardiovascular/pulmonary/skeletal) which would affect your ability to be physically active	<input type="checkbox"/>
Shortness of breath while resting	<input type="checkbox"/>
Concussions	<input type="checkbox"/>
Frequent Fainting / Loss of Consciousness	<input type="checkbox"/>
None of the Above (Not Applicable)	<input type="checkbox"/>

3. Have you noticed a change in difficulty within the last year due to medical (cognitive) issues when participating in the following actions? Please select all that apply.

Table A.2.2: Cognitive Change Questionnaire [5]

Prompt	Yes	No	N/A
Learning how to use unknown tools, appliances, or gadgets	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consistently forgetting the month, day, or year	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Remembering to independently take medication(s)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Keeping up-to-date with important information such as the news	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Getting lost while going for walks or social outings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Thank you for filling out our pre-screening questionnaire.

If you are eligible for participation, we will reach out to you shortly to schedule an in-person appointment. If you have any other questions, please reach out to stoyonag@yorku.ca

A.3 Study Questionnaire

Evaluation of Fitness-Based Recommender System: Pre-Screening Questionnaire

- I consent to participate in this study conducted by Kiemute Oyibo. I have understood the nature of this project and wish to participate. I am not waiving any of my legal rights by signing this form.
 - Yes
 - No
- I understand that my de-identified data will be placed into an open research data repository
 - Yes
 - No
- Do you consent to the use of quotations in any final reports / publications of the research? They will be used to understand the persuasiveness of the generated exercises.
 - Yes
 - No

4. Please indicate your level of agreement to each of the following statements, given the following rating scale: 1 (Strongly Disagree) - 7 (Strongly Agree)

Table A.3.1: Physical Activity Enjoyment Scale (PACES-8) Questionnaire [6]

Prompt	1	2	3	4	5	6	7
Exercise is pleasurable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exercise is not fun	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exercise is very pleasant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exercise is not invigorating	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exercise is very gratifying	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exercise is not exhilarating	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exercise is very stimulating	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exercise is unrefreshing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. Please indicate your level of agreement to each of the following statements, given the following rating scale: 1 (Strongly Disagree) - 5 (Strongly Agree)

Table A.3.2: Privacy Attitudes Questionnaire (PAQ) [7]

Prompt	1	2	3	4	5
I am comfortable giving companies my data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would not like to have my data saved and used to improve the recommender system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Organizations should be allowed to use my information without my consent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would not like the opportunity to delete my data for privacy reasons	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I frequently question why I have to provide personal information to companies	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table A.3.2 (continued)

Prompt	1	2	3	4	5
I do not worry that my input data will be leaked from the recommender system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I value open-source recommender systems over privatized closed-source options	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am uncomfortable with my biomarkers being recorded for recommendation purposes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am comfortable with my personal information (e.g., age, gender) being recorded for recommendation purposes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6. Please indicate your level of agreement to each of the following statements, given the following rating scale: 1 (Strongly Disagree) - 5 (Strongly Agree)

Table A.3.3: Adapted ComTech (HUED-CUP) Questionnaire [1]

Category	Prompt	1	2	3	4	5
Trust	The recommendations appear to be impartial and independent of bias	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trust	The recommendations are not objective	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Expertise	The recommendations came from a knowledgeable source	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Expertise	The recommender system does not have a professional design	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Privacy	The recommender system clearly explains how my data will be used	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Privacy	I feel uncomfortable using the recommender system.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Efficacy	The recommender system told me most of what I needed to know	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Efficacy	The recommendations do not feel as if they were tailored for me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Utility	I find the recommender system useful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table A.3.3 (continued)

Category	Prompt	1	2	3	4	5
Utility	The recommender system has not increased my awareness about exercises and diets	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Usability	The language from the recommender system made the suggestions easy to understand	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Usability	The recommender system was not easy to use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Persuasiveness	The suggested exercises are convincing for me to continue using the recommender system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Persuasiveness	The suggested exercises are relevant for me to continue using the recommender system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Persuasiveness	The suggested exercises would influence me to start or continue using the recommender system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7. Assume you were were planning out your exercises with access to the recommender system. How many minutes of physical activity would you be able to complete within a week?
8. Assuming that you were using the recommender system in a real-life context to support your fitness, how confident are you that you could exercise for at least 150+ minutes per week, given the following scenarios? 1 (No Confidence) - 7 (High Confidence)

Table A.3.4: Self-Efficacy for Exercise Scale (SEE) [8]

Prompt	1	2	3	4	5	6	7
The weather was bothering you	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You were bored by the exercise	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You felt pain when exercising	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You had to exercise alone	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You did not enjoy it	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table A.3.4 (continued)

Prompt	1	2	3	4	5	6	7
You were too busy with other activities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You felt tired	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You felt stressed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You felt depressed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. Please share a bit more about what influenced your above responses. How did your thoughts, experiences, or anything else you'd like us to share affect your final determinations?
10. The following statements reflect how you might approach achieving 150+ minutes of physical activity each week. Please indicate how well each statement relates to you, given the rating scale: 1 (Does Not Describe at All) - 5 (Describes Completely)

Table A.3.5: Adapted Questionnaire from the Exercise Goal-Setting Scale (EGS) and Exercise Planning and Scheduling Scale (EPS) [9]

Prompt	1	2	3	4	5
... I will set a goal to achieve my weekly fitness target	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...I will develop a series of steps to reach my weekly goal.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I will keep track of my progress in meeting my goal.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I will endeavor to achieve the set goal for myself	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I will make my goals public by telling others about it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11. Please share a bit more about what influenced your above responses. How did your thoughts, experiences, or anything else you'd like us to share affect your final determinations?
12. In a few sentences, please share what stood out to you by using the recommender system. What did or did not work well?
13. We are looking for ways to improve the recommendation process by highlighting the most impactful features to change sedentary behaviour. In a few sentences, please share any suggestions or ideas that could help make the fitness-based recommender system better for you or others.

14. How did you learn about the user study?

- LinkedIn
- Discord
- Reddit
- E-mail (grads@eecs.yorku.ca)
- Poster
- Word of Mouth

15. What is your gender?

- Male
- Female
- Non-Binary
- Prefer not to say

16. Please provide us with your age range.

- 18 - 19
- 20 -22
- 23 - 25
- 26 - 28
- 29 - 30
- Prefer not to answer

17. What is the highest level of education you've completed, or are in the process of completing?

- Highschool
- Vocational College
- Undergraduate
- Masters
- PhD
- Post Doc
- Prefer not to answer

18. What is/was your major (and minor, if applicable)?

19. What is your ethnicity?

- White or Caucasian
- Arab
- Black (e.g., African, American, Canadian Caribbean)
- Chinese
- Japanese
- Korean
- Filipino
- Non-White Latin American
- Non-White West Asian (e.g., Iranian, Lebanese, Afghan)
- South Asian / East Indian (e.g., Bangladeshi, Pakistani, Indian, Sri Lankan)
- South East Asian (e.g., Burmese, Cambodian, Laotian, Malaysian, Vietnamese)
- Prefer not to answer

20. Please enter your approximate weight (in kg). If you prefer not to answer, please enter 0.

21. Please enter your approximate height (in m). If you prefer not to answer, please enter 0.

22. On average, how many minutes per week do you participate in moderate to rigorous physical activity? Please describe the activities if applicable (e.g., cycling).

Thank you for your time!

Questions About the Research?

If you have questions about the research in general or about your role in the study, please feel free to contact Kiemute Oyibo by email (kiemute.oyibo@yorku.ca). You may also reach out to the Electrical Engineering and Computer Science (EECS) Graduate program by telephone at (416)-650-8215, or by email (gpa@eecs.yorku.ca). This research has received ethics review and approval by the Human Participants Review Sub-Committee, York University's Ethics Review Board and conforms to the standards of the Canadian Tri-Council Research Ethics guidelines. If you have any questions about this process, or about your rights as a participant in the study, please contact the Manager, Research Ethics in the Office of Research Ethics, York University (e-mail ore@yorku.ca). This office oversees the ethical conduct of research studies and is not part of the study team. Everything that you discuss will be kept confidential.

B Datasets

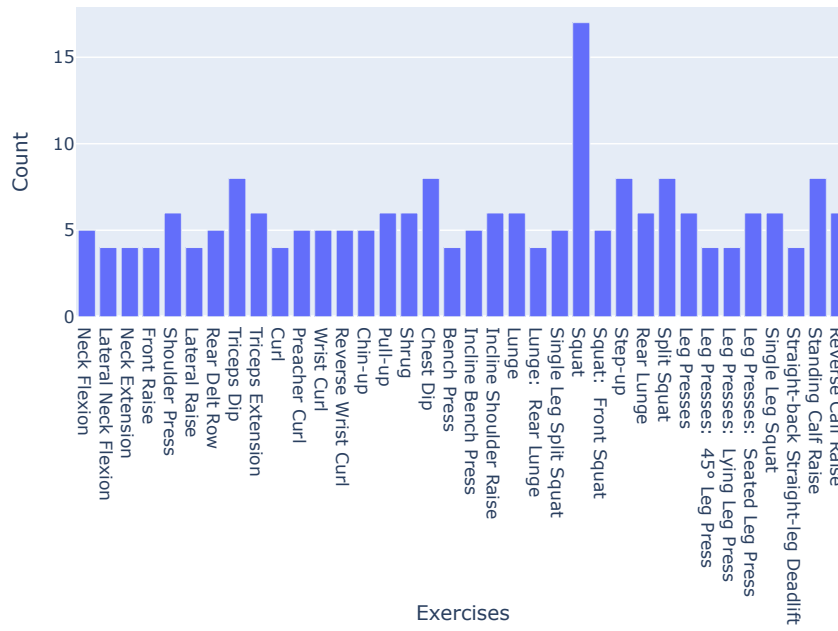


Figure B.1: Murarka's Dataset [3]: Visualization After Removing Labels with less than 4 Occurrences

C Prompts

```
1 template = """You are a friendly bariatric healthcare professional. You have the
   following information about the patient:
2
3 Age: {age}
4 Gender: {gender}
5 Weight (kilograms): {weight}
6 Height (meters): {height}
7 BMI: {bmi}
8 Health Conditions: {health}
9 Social Preference: {social}
10 City: {city}
11 Location Preference: {location}
12
13 Using this information, answer their question with
14 actionable examples and explanations.
15
16 Question: {question}
17
18 Answer:
19 """
```

Figure C.1: Prompt Template: Context-Aware Recommender System

```
1 question = "Given my condition, how can I lose weight? Provide me with exercises and
   dietary recommendations."
```

Figure C.2: Control Prompt: Context-Aware Recommender System

```
1 question = f'Given my condition, how can I lose weight? Provide me with exercises and
   dietary recommendations. Use the following persuasive strategies in your answer: {
   asyncio.run(query_frameworks('persuasive design strategies'))}'
```

Figure C.3: Experimental Prompt: Context-Aware Recommender System

```
1 template = """Explain why the following exercise is beneficial to perform.  
2 Make your answer persuasive. Suggest how to incorporate it into a weekly schedule:  
3 To make the answer persuasive, consider motivation, ability, and triggers (Fogg  
   Behavioural Model).  
4  
5 Exercise(s): {rs_output}  
6  
7 Answer:  
8 """  
9  
10 prompt = PromptTemplate.from_template(template)  
11 return llm.invoke(prompt.invoke(rs_output)).replace("</s>", "")
```

Figure C.4: Experimental Prompt: Content-based Recommender System

D Application (Screenshots & Instruments)

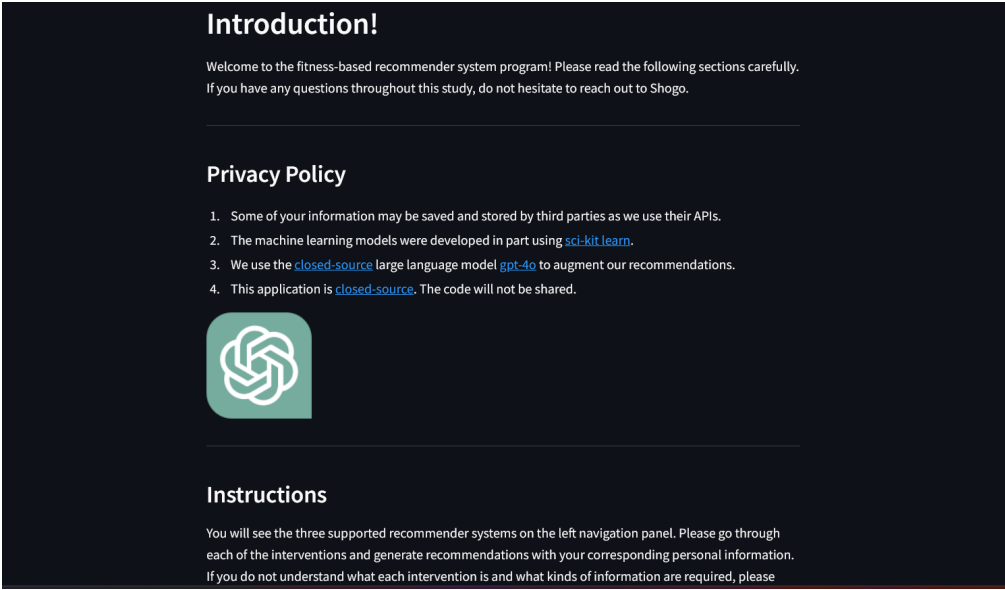


Figure D.1: Control Group: Introduction Page (Privacy Policy)

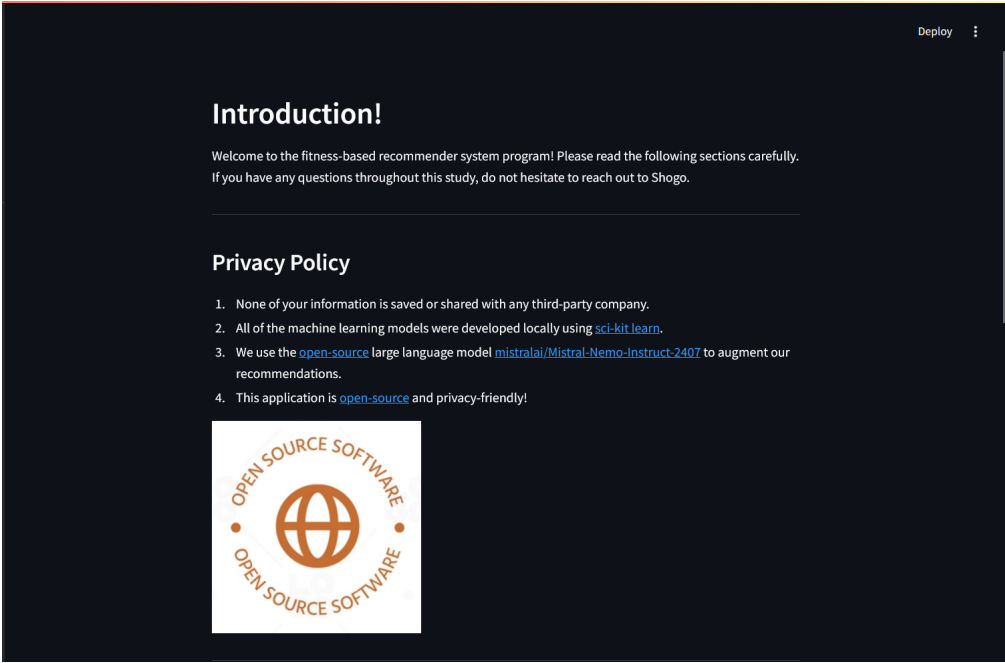


Figure D.2: Experimental Group: Introduction Page (Privacy Policy)

```
> Entering new AgentExecutor chain...
  I need to provide a weight loss plan that is tailored to this individual's health conditions, social preferences, and location preference.
Action: Search
Action Input: "weight loss plan for depression and anxiety, individual activity, indoors"
Observation: Just 30 minutes of brisk walking each day can burn around 150 calories , making various walking exercises a valuable component of any weight loss plan. The indoor walking challenge offers a comprehensive full-body workout that incorporates cardio exercises, engaging multiple muscle groups and joints while boosting metabolic rate to promote ... 2. The 28-day indoor walking challenge boosts your mood. Physical activity isn't just good for your body; it's great for your mind, too. "The mental health benefits of walking are huge," says Stasia Patwell, certified fitness trainer. "Walking can boost your mood, lower anxiety, help with stress management and improve your sleep." Many individuals find themselves caught in a complex web of depression, anxiety, and weight management issues, struggling to find a path towards overall health and happiness. This comprehensive guide aims to shed light on the intricate relationship between mental health and weight loss, offering practical strategies and insights to help you ... But even small amounts of physical activity can be helpful. Being active for short periods of time, such as 10 to 15 minutes at a time, throughout the day can add up and have health benefits. Regular exercise may improve depression or anxiety symptoms enough to make a big difference. That big difference can help kick-start further improvements. Indoor Activities for Seniors: Simple Ways to Stay Fit and Active for Quality of Life ... : Exercise, including walking, releases endorphins which can improve mood, reduce stress, and alleviate symptoms of depression and anxiety (4). ... Define what you want to achieve with your indoor walking plan, whether it's weight loss, improved fitness ...
Thought: I also need to provide dietary recommendations.
Action: Search
Action Input: "dietary recommendations for depression and anxiety"
```

Figure D.3: Experimental Group: Verbose Mode of Context-Aware Agent

E Additional Results

E.1 Participant Demographics

Table E.1: Participant Demographics: Age Groups

Age	Gender		Total
	Male	Female	
18 - 19	5	6	11
20 - 22	18	14	32
23 - 25	10	5	15
26 - 28	4	5	9
29 - 30	0	5	5
Prefer not to answer	1	0	1
Total	38	35	73

Table E.2: Participant Demographics: Ethnicity

Ethnicity	Gender		Total
	Male	Female	
White or Caucasian	5	6	11
Arab	1	0	1
Black (e.g., African, American, Canadian Caribbean)	4	2	6
Chinese	5	2	7
Japanese	0	0	0
Korean	0	1	1
Filipino	0	2	2
Non-White Latin American	4	0	4
Non-White West Asian (e.g., Iranian, Lebanese, Afghan)	3	5	8
South Asian / East Indian (e.g., Bangladeshi, Pakistani, Indian, Sri Lankan)	15	8	23
South East Asian (e.g., Burmese, Cambodian, Laotian, Malaysian, Vietnamese)	0	5	5
Prefer not to answer	1	4	5
Total	38	35	73

Table E.3: Participant Demographics: Education Level and Major of Study

Study	Education	Gender		Total
		Male	Female	
Computer Science	Undergraduate	9	3	12
	Masters	4	3	7
	PhD	0	0	0
	Vocational College	0	0	0
	Total	13	6	19
Mechanical Engineering	Undergraduate	1	0	1
	Masters	0	0	0
	PhD	0	1	1
	Vocational College	0	0	0
	Total	1	1	2
Biomedical Science	Undergraduate	0	4	4
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	0	4	4
Robotics Engineering	Undergraduate	0	0	0
	Masters	0	1	1
	PhD	0	0	0
	Vocational College	0	0	0
	Total	0	1	1
Computer Engineering	Undergraduate	4	0	4
	Masters	1	0	1
	PhD	1	0	1
	Vocational College	0	0	0
	Total	6	0	6
Philosophy	Undergraduate	1	0	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	1	0	1

Table E.3 (continued)

Study	Education	Gender		Total
		Male	Female	
Space Engineering	Undergraduate	1	0	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	1	0	1
Cyber Security	Undergraduate	1	0	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	1	0	1
Applied Mathematics	Undergraduate	1	0	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	1	0	1
Kinesiology	Undergraduate	2	3	5
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	2	3	5
Fashion & Design	Undergraduate	0	0	0
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	1	0	1
	Total	1	0	1
Neuroscience	Undergraduate	2	6	8
	Masters	1	1	2
	PhD	0	0	0
	Vocational College	0	0	0
	Total	3	7	10

Table E.3 (continued)

Study	Education	Gender		Total
		Male	Female	
Software Engineering	Undergraduate	2	0	2
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	2	0	2
Biology	Undergraduate	0	0	0
	Masters	0	1	1
	PhD	0	0	0
	Vocational College	0	0	0
	Total	0	1	1
Psychology	Undergraduate	0	4	4
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	0	4	4
Spanish and French	Undergraduate	0	1	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	0	1	1
Theatre	Undergraduate	0	1	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	0	1	1
Visual Arts	Undergraduate	0	0	0
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	1	1

Table E.3 (continued)

Study	Education	Gender		Total
		Male	Female	
	Total	0	1	1
Global Health	Undergraduate	0	1	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	0	1	1
Biotechnology	Undergraduate	0	1	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	0	1	1
Environmental Design	Undergraduate	0	1	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	0	1	1
Transportation Engineering	Undergraduate	0	0	0
	Masters	0	1	1
	PhD	0	0	0
	Vocational College	0	0	0
	Total	0	1	1
Electrical Engineering	Undergraduate	2	0	2
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	2	0	2
Physics	Undergraduate	1	0	1
	Masters	0	0	0
	PhD	0	0	0

Table E.3 (continued)

Study	Education	Gender		Total
		Male	Female	
	Vocational College	0	0	0
	Total	1	0	1
Chemistry	Undergraduate	1	0	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	1	0	1
Civil Engineering	Undergraduate	1	0	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	1	0	1
Earth and Atmospheric Science	Undergraduate	0	1	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	0	1	1
Business Economics	Undergraduate	1	0	1
	Masters	0	0	0
	PhD	0	0	0
	Vocational College	0	0	0
	Total	1	0	1
Total	Undergraduate	30	26	56
	Masters	6	7	13
	PhD	1	1	2
	Vocational College	1	1	2
	Total	38	35	73