

[BITE or BYTE]: Developing and Evaluating An Educational Game For Teaching AI Literacy in Nutritional Contexts

Roberta Pošiūnaite



4th Year Project Report
Computer Science
School of Informatics
University of Edinburgh
2025

Abstract

This study explores the intersection of generative AI (GenAI), nutrition misinformation, and game-based learning by developing and evaluating an educational game that aims to teach users critical evaluation skills of AI-generated nutritional content. As GenAI chatbots become increasingly accessible, users may encounter potentially harmful nutritional information that contains inaccuracies or AI hallucinations. Through an iterative design process involving focus groups and user testing, I developed an education game where players distinguish between AI-generated and written and literature-reviewed nutritional responses while assuming different dietary personas. Qualitative analysis revealed that participants regularly use AI chatbots but remain skeptical of their responses, especially in nutritional contexts. After gameplay, participants described a loss of trust in AI-generated content, especially for nutritional content, and improved AI literacy skills. This study demonstrates the potential of game-based learning in developing critical AI evaluation skills, particularly in sensitive domains such as nutrition, while highlighting the need for further development of the system.

Research Ethics Approval

This project obtained approval from the Informatics Research Ethics committee.

Ethics application number: 774516

Date when approval was obtained: 2024-11-28

The participants' information sheet and a consent form are included in the appendix.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(Roberta Pošiūnaitė)

Acknowledgements

Thank you to my supervisor, Dr. Tara Capel, for her invaluable guidance and unwavering support throughout this project. Her feedback and encouragement kept me motivated and excited to complete this project.

I would like to thank the participants of the study for taking the time to help me with this project, for their feedback, and for speaking clearly enough that fixing the transcripts did not take too long.

My deepest appreciation goes to my friends and family for their endless support and constant reassurance.

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Objectives	2
1.3	Approach and Report Structure	2
2	Background Chapter	4
2.1	Generative AI in Wellbeing, Health and Nutrition	4
2.2	AI Hallucinations	6
2.3	AI-Generated Content in Online Spaces	8
2.4	Game-Based Learning and Serious Games	8
2.5	Conclusion and Research Direction	9
3	Methods	11
3.1	Research Design	11
3.2	Participants	11
3.3	Session 1: GenAI Usage and Prototype Evaluation	13
3.3.1	Semi-Structured Focus Group	13
3.3.2	Prototype Evaluation and Discussion	13
3.4	Session 2: Game Evaluation	14
3.4.1	Think-Aloud Study	14
3.4.2	Group Discussion	14
3.5	Data Analysis	15
3.5.1	Thematic Analysis	15
3.5.2	Pain Point Identification	15
4	Focus Group 1: GenAI Usage and Prototype Evaluation	16
4.1	Discussions of Nutrition Misinformation	16
4.2	Use of Generative AI Tools	17
4.3	Nutrition and Generative AI	18
4.4	Prototype Evaluation and Suggestions for Improvement	18
4.4.1	Persona Creation	18
4.4.2	AI-Generated vs Literature-Reviewed Response Highlighting	18
4.4.3	Answer Selection	19
4.4.4	Response Interface	19
4.4.5	Explanation Pop-Ups	19
4.4.6	Countdown and Points	19

4.4.7	Number of Questions and End Goal	19
4.4.8	Question Difficulty	19
4.5	Key Takeaways and Implications for Game Design	20
5	Game Design and Development	21
5.1	Technology Stack	21
5.2	Game Theme and Overall Design	21
5.3	Interface Features and Game Mechanics	22
5.3.1	Before Gameplay	22
5.3.2	During Gameplay	23
5.3.3	After Gameplay	24
5.3.4	Throughout Gameplay	25
6	Focus Group 2: Functioning Prototype Evaluation	29
6.1	Game Design and Mechanics Feedback	29
6.1.1	Landing Page	29
6.1.2	Username Generation	29
6.1.3	Persona Creation	30
6.1.4	Instructions	30
6.1.5	Game	30
6.1.6	Leaderboard	31
6.1.7	Other Comments	31
6.2	Post-Play Discussion	32
6.3	Game Improvements	32
7	Discussion	34
7.1	GenAI for Nutritional Advice	34
7.2	Effectiveness of Game-Based Learning for AI Literacy	35
7.3	Limitations	36
7.4	Further Research	36
8	Conclusions	38
	Bibliography	39
A	Supporting Ethics Documents	44
A.1	Participant Information Sheet	44
A.2	Participant Consent Form	44
B	Initial Figma Prototype	50
C	Post-Focus Group 1 Functioning Prototype	57
D	Final Functioning Prototype	69
E	Analysis Across Focus Groups	72
E.1	Focus Group 1 Group Discussion Thematic Analysis	72

E.2 Focus Group 2 Group Discussion and Individual Think-Aloud Thematic Analysis 73

E.3 Focus Group 2 Pain Point Identification 73

Chapter 1

Introduction

1.1 Motivation

In November 2022, OpenAI released its generative AI (GenAI) chatbot, ChatGPT, to the public [37]. Within two months, it was estimated that the application had garnered 100 million users, becoming the fastest-growing consumer application in history [22]. In the years that have followed, many other GenAI chatbots have entered circulation: Anthropic’s Claude, Google’s Gemini (previously Bard), Microsoft’s Copilot, DeepSeek, and many more [2] [11] [32] [12]. These chatbots are becoming a tool used in everyday life, but the wide adoption of such technology raises questions about the critical evaluation of AI-generated content, particularly in contexts where misinformation can lead to detrimental health outcomes.

One such context is nutrition. Large Language Models (LLMs) use machine learning to generate human language, trained on large amounts of data to recognize patterns and produce responses [50]. LLMs such as ChatGPT can produce “hallucinations” - information that, although stated confidently, is factually incorrect and can pose serious risks to users in the domain of nutrition or health [23]. Even if users do not use an AI chatbot for nutritional advice, they may encounter AI-generated nutritional content online. Poor quality and inaccurate nutritional information has been widespread on websites and social media even before the introduction of GenAI chatbots [13]. This existing environment of misinformation provides grounds for potentially inaccurate AI-generated content to spread. As AI is universally adopted, publications, websites, and individuals may use these tools to generate potentially incorrect nutritional advice or information and disseminate it online, further amplifying the misinformation landscape.

Recent research has started to explore the dangers of AI-generated nutritional content. For example, Niszczoła and Rybicka [35] demonstrated that ChatGPT may include allergens in meal plans for users with explicitly stated allergies, Haman et al. [19] showcased how ChatGPT might miscalculate macronutrients in foods, and Ponzo et al. [40] found that AI chatbots may provide contradictory dietary advice for those with multiple non-communicable diseases. These studies highlight the limitations of GenAI chatbots in handling complex nutritional scenarios and underscore the prevalence of AI hallucinations.

Despite the risks, users are still likely to employ GenAI chatbots for nutritional advice and may encounter AI-generated dietary guidance online. This trend highlights a need to inform users of the risks of using AI tools in nutritional contexts and teach them critical evaluation skills and ways to discern AI-generated content. Game-based learning offers a promising approach to address this need, as it has been found to successfully teach critical thinking skills and has been employed to teach users about the limitations of AI [41] [14].

This research seeks to develop and evaluate a game-based educational tool that teaches users about the potential risks of using AI chatbots for nutritional advice and critical AI literacy and evaluation skills to discern AI-generated content in a safe environment. Inspired by the work of Kirk et al. [26], where participants assessed the correctness, actionability, and comprehensibility of AI-generated responses to nutritional questions by comparing them to the responses written by nutritionists, I aim to employ the mechanism of comparison to enhance users' AI literacy and promote safer engagement with AI technologies in nutritional contexts.

1.2 Objectives

This project aims to develop and evaluate an educational game that enhances users' critical literacy skills in distinguishing AI-generated content and promotes safer engagement with GenAI chatbots in nutrition-related contexts. The specific objectives are:

- To investigate users' current interactions with GenAI tools, their trust in AI-generated content, and their use of AI chatbots for nutritional information.
- To design and develop a game-based educational tool that teaches users to identify and evaluate AI-generated nutritional content, with a particular focus on recognizing hallucinations, thus exploring the potential risks of GenAI in nutritional and health-based domains.
- To evaluate the game in its usability, ability to alter users' trust levels in AI-generated nutritional content, and enhancement of users' critical evaluation skills of AI-generated content.

1.3 Approach and Report Structure

The study incorporates qualitative research methods to develop and evaluate the game-based educational tool. The research follows an iterative process, which involves focus groups, prototype development, user testing, and thematic analysis, all detailed in the report. Chapter 2: Background details a literature review examining four areas: the current state of GenAI in well-being, health and nutritional contexts, AI hallucinations and their implications for nutritional information, the spread of AI-generated content in online spaces, and the effectiveness of game-based learning and serious games. Chapter 3: Methods, outlines the research design. Chapter 4: Focus Group 1 presents the findings from the initial focus group exploring participants' experiences with GenAI,

nutritional information, the intersection of the two, and their evaluation of the initial game prototype. In Chapter 5: Game Design and Development, details of the technical implementation and design decisions for the game can be found. Chapter 6: Focus Group 2 presents the evaluation of the functioning prototype through individual think-aloud sessions and group discussions. This is followed by Chapter 7: Discussion, which examines the findings and limitations of the study, as well as directions for future research. Finally, Chapter 8: Conclusions summarizes the key findings of this research.

Chapter 2

Background Chapter

Generative Artificial Intelligence (GenAI) is becoming increasingly more accessible, with users employing it to receive answers to simple questions or to complete more complex tasks like generating images and videos. The intersection of GenAI and nutrition has introduced new opportunities and challenges in health and wellness, raising important questions about AI hallucinations and the generation of incorrect and potentially harmful dietary information. This chapter explores the current landscape of GenAI tools in nutritional advice and wellness while examining the challenges and limitations that exist in the field, focusing on AI hallucinations. We will also explore game-based learning and serious games as potential approaches to educate users about the responsible use of GenAI in nutritional contexts and critical evaluation skills of AI-generated content.

2.1 Generative AI in Wellbeing, Health and Nutrition

As GenAI models like ChatGPT grow in popularity, research has begun to investigate their potential and limitations in the field of nutrition. Recent studies have examined the accuracy and safety of GenAI chatbots in nutritional contexts.

Tsiantis et al. [49] performed a systematic literature review to explore the scope of research conducted at the intersection between AI and nutrition. They searched through academic databases for relevant publications since 2022, including ten articles that met their inclusion criteria [49]. The articles were categorized into five areas: nutritional content analysis, diet planning, meal variability assessment, dietetics practice and education evaluation, and food appropriateness determination [49]. Their findings highlight the strengths of ChatGPT in generating meal plans with nutritional accuracy, adapting dietary guidelines, and providing reliable answers to common nutrition questions [49]. However, many limitations were also identified, such as inaccuracies in food quantities, the risk of including unwanted allergens, and the use of outdated information [49].

ChatGPT has been shown to present reasonably accurate nutritional information in an accessible way for straightforward questions but its performance deteriorates when handling more complex scenarios, with it including allergens in meal plans and contra-

dicting clinical guidelines [26] [40] [35]. Kirk et al. [26] carried out a study comparing ChatGPT's responses to common nutrition questions and answers provided by real dietitians. Other dietitians and experts in the fields graded the answers based on several factors: scientific correctness, actionability, and comprehensibility [26]. Similarly, focusing on non-communicable diseases (NCDs), Ponzo et al. [40] examined the appropriateness of ChatGPT's dietary advice in contrast to international guidelines for several NCDs. Non-communicable diseases, also known as chronic diseases, tend to be of long duration and are the result of a combination of genetic, physiological, environmental, and behavioral factors [38]. Ponzo et al. [40] prompted ChatGPT with questions related to seven NCDs: dyslipidemia, arterial hypertension, type 2 diabetes mellitus, obesity, non-alcoholic fatty liver disease (NAFLD), chronic kidney disease (CKD), and sarcopenia and tested the chatbot's answers by comparing them to recommendations from clinical guidelines. The prompts were formulated to mimic how patients would enquire healthcare professionals [40]. To assess the credibility of ChatGPT's nutritional recommendations for those with food allergies, Niszczoła and Rybicka [35] investigated whether ChatGPT can reliably create safe and accurate meal plans that would exclude allergens but still ensure that nutritional needs are met. The researchers created a persona of a 30-year-old woman with specific food allergies and designed prompts for ChatGPT to generate diets at four levels of restriction (A-D) [35]. Each level of restriction considered a different food allergen and additional goals such as calorie targets or another allergen [35]. The plans generated by ChatGPT were then analyzed and evaluated on their safety, accuracy, and attractiveness by a qualified dietitian [35]. Kirk et al. [26] found that ChatGPT consistently received higher grades than human dietitians in providing overall quality answers. The study suggested that ChatGPT can effectively respond to and communicate complex nutritional information in a way that is understandable to the user [26]. Ponzo et al. [40] found that ChatGPT provided generally accurate and clear advice, showcased by appropriateness rates ranging from 55.5% for sarcopenia and 73.3% for NAFLD. However, ChatGPT's ability to provide accurate advice diminished when more complex queries were presented that required more customized strategies, such as when a patient has several NCDs [40]. They also found a few recommendations to be contradictory to the guidelines [40]. Niszczoła and Rybicka [35] found that while ChatGPT successfully excluded allergens in most cases, it made errors as well, including the inclusion of almond milk in nut-free diets. It also struggled to calculate energy values and food quantities, especially when specific calorie targets were set [35]. ChatGPT was also inclined to generate monotonous diets with repetitive ingredients and meals, which made the diets unattractive [35]. The inclusion of allergens in a meal plan, as demonstrated by Niszczoła and Rybicka [35], raises a serious safety issue, as including allergens could potentially cause severe reactions in allergic individuals. The research suggests that although ChatGPT may be a useful tool as a supplementary resource, it should not and cannot replace human dietitians.

One study showcases how ChatGPT may be able to provide generally accurate approximations of calorie content consistently, but not necessarily for different micronutrients [19]. Haman et al. [19] compared the nutritional values provided by ChatGPT for 236 different food items against data from the United States Department of Agriculture (USDA) Food Data Central. They analyzed the accuracy of ChatGPT's estimations for energy, carbohydrates, protein, lipids, and water by repeating each query five times

and in new chats and evaluating the responses by comparing them to the reference values from the USDA database [19]. Haman et al. [19] found that 97% of ChatGPT's estimations fell within a 40% difference of the USDA values for energy but that the accuracy varied for different nutrients. For example, only about 70% of estimations of total lipids (fat) were within a 40% difference of the USDA values, showcasing a lack of accuracy by ChatGPT [19]. A relatively low coefficient of variation indicated that ChatGPT displayed consistency in its responses [19]. Miscalculations of macronutrients can lead to overconsumption or underconsumption of essential nutrients, which can be dangerous for users.

These studies showcase the promising capabilities and significant limitations of GenAI chatbots such as ChatGPT in nutritional contexts. While ChatGPT demonstrates an ability to generate meal plans with nutritional accuracy, adapt dietary guidelines, and provide accessible answers to common nutrition questions, it also suffers from several critical flaws, lacking precision and safety guarantees, which can harm a user utilizing ChatGPT for nutritional guidance. Many people may still rely on GenAI chatbots for nutritional advice, making it essential to educate users about the potential risks and teach them to evaluate AI-generated responses more critically.

2.2 AI Hallucinations

GenAI is good at predicting what word should go next in a sentence - that is how it sounds fluent, but just because it sounds good does not mean it's true [23]. Waldo and Boussard [51] illustrate this point by saying: "Unlike the philosophical dictum that the sentence "Grass is green" is true because, in the real world, grass is green, a GPT will tell us that grass is green because the words "grass is" are most commonly followed by "green." It has nothing to do with the color of the lawn." The predictability of what comes next in a sentence may generate low-quality text [23]. Therefore, randomness is incorporated into decoding [23]. Randomness is the utilization of stochastic processes to generate diverse text [6]. Although randomness helps generate better-sounding text, it has been "positively correlated with an increased risk of hallucinations" [23]. In natural language processing (NLP), a hallucination is typically defined as a phenomenon where "the generated content appears nonsensical or unfaithful to the provided source content" [23]. For Large Language Models (LLMs), Huang et al. [23] propose defining two primary types of hallucinations: factuality hallucination and faithfulness hallucination. Factuality hallucination emphasizes the "discrepancy between generated content and verifiable real-world facts, typically manifesting as factual inconsistencies" [23]. Whereas faithfulness hallucination "captures the divergence of generated content from user input or the lack of self-consistency within the generated content" [23]. Although reasons for LLM hallucinations are not fully understood [10], Huang et al. [23] suggest several other factors that could contribute to the hallucinated information presented by LLMs. One such factor can be caused by the data used when training the models [23]. LLMs are trained on two main components: pre-training data, through which an LLM will acquire its factual knowledge, and alignment data, which will train an LLM to follow user instructions [23]. If the pre-training data is flawed and plagued with biases or misinformation, it can lead to hallucinations by the LLM [23]. For

example, although biases and hallucinations differ, some biases are closely linked to hallucinations, particularly those related to gender and nationality [23]. An LLM might associate nursing with women, even though gender is not explicitly mentioned, illustrating context inconsistency [23].

These hallucinations may harm users when an LLM is used for nutritional information. Mbakwe et al. [31] argue that ChatGPT's training on potentially biased internet medical content can skew the AI chatbot's knowledge. Mbakwe et al. [31] suggest that the sources of bias in medical knowledge that could be present in ChatGPT's training data could be from research performed in high-income countries, which can lead to a lack of representation of health conditions, textbooks with non-representative studies which may not accurately reflect the global population and dominance of content from certain high-powered academic institutions which may overlook other findings. Mbakwe et al. [31] also point out that "those at the table with the loudest voice, which in this case produces the content that dominates the internet, will shape the input, and therefore the output of LLMs," highlighting how if certain perspectives are overrepresented online due to other factors and not their validity, an LLM will likely reflect this imbalance.

A systemic review by Denniss et al. [13] identified content analysis studies "published in English after 1989 that evaluated the quality and/or accuracy of nutrition-related information published on websites or social media." They created a coding framework that classified the studies' findings about information quality and/or accuracy as poor, good, moderate, or varied [13]. Denniss et al. [13] found "that of the studies that evaluated nutritional information quality on websites and social media, 47.1% and 62.5% were classified as poor, respectively." Similarly, they found "that between studies that evaluated accuracy on websites and social media, with 47.7% and 50% classified as poor, respectively" [13]. This highlights the high level of poor quality and inaccurate nutrition-related information on websites and social media. If LLMs are trained on this inaccurate and poor-quality information, they are likely to hallucinate inaccuracies or provide wrong information in their responses to user queries about nutrition.

Even if a user requests for sources of the information provided by an LLM, the sources may be hallucinated themselves. Chelli et al. [9] illustrate the rates of hallucination and reference accuracy of ChatGPT and Bard (now Gemini) for systemic reviews. Chelli et al. [9] compared LLM performance to human-conducted systemic reviews in the context of scientific writing, using shoulder rotator cuff pathology as a gold standard. To calculate hallucination rates, they considered papers "hallucinated" if any of the two following information were wrong: title, first author, or year of publication [9]. Chelli et al. [9] found that both ChatGPT and Bard exhibited high hallucination rates, with GPT-3.5 and GPT-4 having hallucination rates of 39.6% and 28.6%, respectively. This undermines a common verification strategy of requesting sources since the sources themselves may be fabricated or inaccurately cited, potentially misleading users who don't verify the LLM-produced citations.

The research highlights how LLMs may provide nutritional information that sounds confident but could be inaccurate due to hallucinations, as the pre-training data for LLMs likely includes a significant amount of poor-quality nutritional information from websites and social media. It is essential to ensure users do not rely on LLM-generated

information when making critical health or nutrition decisions, as these models cannot guarantee factual accuracy in this domain.

2.3 AI-Generated Content in Online Spaces

As AI tools become increasingly accessible, AI-generated content is spreading across online spaces, including news articles, blogs, and social media posts, raising significant concerns about information quality [1]. Recent studies predict an alarming increase in AI-generated or AI-translated content online in the coming years [16] [47]. This trend creates a troubling feedback loop - as AI-generated content becomes widespread, LLMs may increasingly train on content produced by other LLMs rather than human-written text [44]. This can lead to a model collapse - “a degenerative process whereby, over time, models forget the true underlying data distribution” [30], and a decrease in AI output quality, leading to more hallucinated and incorrect information being generated and spread [45]. Given that online spaces are already polluted by poor-quality and inaccurate nutritional information [13], the addition of AI-hallucinated nutritional content poses substantial risks to users. Therefore, individuals need to develop skills to identify AI-generated text, enabling them to approach such information critically even when encountering it outside dedicated AI platforms.

2.4 Game-Based Learning and Serious Games

A game is “defined as a physical and/or mental contest that is played according to specific rules, with the sole goal of amusing or entertaining the participant(s)” [28]. Plass et al. [39] argue that any gameplay with defined learning outcomes corresponds to game-based learning (GBL). Games can capture the attention of a learner through compelling narratives, challenges, and rewards, fostering a sense of investment in the learning process [20]. Furthermore, GBL enhances retention and recall by providing experiential learning opportunities, allowing learners to apply knowledge in realistic scenarios [25]. GBL can help learners develop problem-solving and critical-thinking skills. Shaffer [43] argues that educational games require players to analyze situations, make strategic decisions, and adapt to new information, all of which contribute to higher-order thinking. Some research has been conducted to understand how game-based learning can be used to teach 21st-century skills, such as critical thinking, creativity, collaboration, and communication [41]. Qian and Clark [41] examined literature in “regard to game-based learning and identified 29 studies which targeted 21st-century skills as outcomes.” They performed a systematic literature review of studies published between 2010 and 2014, from which they found that in the 29 investigated papers, the effects of GBL on 21st-century skills reported “a total of 97 outcomes, with 85% of the results being significant (i.e., $p < 0.05$), and roughly one-third (34%) being associated with medium to large effect sizes” [41]. These results highlight the possibility of employing game-based learning methodologies and frameworks to teach users skills in critical thinking and communication. There are concerns surrounding GBL; some argue that games are “just another technological fad, which emphasizes superficial learning” [24]. Emes [15] also argues that there is no relationship between game-based learning

and high academic achievements or psychopathology.

Where game-based learning is an approach to teaching in educational contexts, serious games are a tool used in the learning process [5] [42]. Game-based learning can be achieved by using any games in an educational setting, even ones not designed for educational purposes [46]. Serious games are designed specifically for purposes other than mere entertainment [46]. Serious games do not have to be used in a formal educational setting - they can be leveraged for professional development, training, simulation, and informal learning [5] [42]. Schrader [42] argues that both serious games and game-based learning take advantage of the fundamental entertaining characteristics to produce goals, outcomes, and experiences.

For Safer Internet Day 2025, Microsoft announced the release of “CyberSade AI: Dig Deeper,” an education game in Minecraft that “focuses on the responsible use of AI” [33]. In a blog post about the learning experience, Minecraft Education explains how the game leads students to learn about AI hallucinations and challenges surrounding data privacy and information integrity when it comes to AI, by having users play through made-up scenarios within the game [14]. Arai et al. [3] have designed a game for education in AI security leveraging episodic memory. They evaluated the game with a questionnaire survey of 48 non-expert users and found that “74% of users considered the game scenarios effective, and 81% considered game functionalities effective” [3]. In May of 2023, Caroline Buttet and Emmanuel Durgoni, Artists in Residence at Google Arts & Culture Lab released a game experiment called “Odd One Out”, in which the player is presented with four images, three of which are real pieces of art hosted at Google Arts and Culture, and one is created by Google AI [18]. The aim is for the player to find the AI-generated art [18]. This game showcases the troubling speed at which AI-generated art is improving, as it is often indistinguishable from real art. These games highlight a growing interest in building games to educate people on the dangers and responsible use of GenAI tools. Game-based learning approaches can be valuable when teaching critical evaluation of AI-generated nutritional advice. A gamified approach could engage learners through competition and reward systems but also create a safe environment to practice critical approaches to AI-generated content. A serious game about GenAI and nutrition could teach users how to approach AI-generated nutritional content critically, recognize its limitations, and understand the potential risks of relying on AI for dietary guidance.

2.5 Conclusion and Research Direction

I have examined four interconnected areas: the current state of GenAI in nutritional contexts, the presence and impact of AI hallucinations, the abundance of AI-generated content in online spaces, and the effectiveness of game-based learning for critical thinking development. The research discussed showcases that while GenAI chatbots like ChatGPT can provide accurate nutritional information for simple queries, they have significant limitations when handling more complex cases, such as providing meal plans for individuals with allergies, calculating macronutrients, and answering queries for individuals with multiple NCDs [40] [35] [19]. The generation of this inaccurate information may stem from AI hallucinations, which are particularly concerning in

health contexts where pre-training data may include poor-quality or inaccurate nutritional information from websites and social media [23] [13]. Some game-based educational approaches to teaching AI literacy have emerged [14] [3] [18], but to the best of my knowledge, none have had a focus on nutrition. In this project, I aim to address this gap by developing a game-based educational tool inspired by Kirk et al.'s [26] work comparing AI-generated responses with human-written content to queries about nutrition. However, rather than evaluating the accuracy and comprehensibility of the AI's responses, this project leverages the comparison mechanism to actively teach users critical evaluation skills and raise awareness about the potential dangers of relying on AI tools for nutritional advice.

Chapter 3

Methods

3.1 Research Design

I employed a qualitative research design to explore the development of an educational game aimed at promoting critical thinking about generative AI (GenAI) in nutrition. The research follows a user-centered approach, incorporating principles of Human-Computer Interaction (HCI), such as Nielsen’s 10 Heuristics [34] and the Gestalt Principles [27], to assess usability and interactions. Data was collected through focus groups, prototype evaluations, and user testing, employing thematic analysis and pain point identification. I chose a qualitative approach as it allows an in-depth exploration of user experiences, perceptions, and interactions with AI-generated content. Focus groups and prototype evaluations provide valuable contextual insights that would be difficult to capture using quantitative methods. This approach enabled me to capture nuanced user responses to AI-generated nutritional content, revealing what users found challenging to identify and why certain AI outputs are more convincing than others. The methods I employed allowed different themes to emerge and guide the iterative development of the game mechanics and the educational content.

3.2 Participants

I recruited four participants for the study, with a mix of backgrounds and experiences. I recruited the participants by asking people I knew about their experiences with GenAI, nutritional information online, and using GenAI for nutritional advice. I selected people who had varying experiences in these domains. A small sample size would mean individual opinions and preferences would have a bigger impact and could lead to an inability to generalize results and potential bias in feedback. However, the small sample size also enabled in-depth qualitative analysis, facilitating a comprehensive understanding of user interactions and perceptions. The study required detailed feedback rather than generalizability. Thus, a smaller group allowed me to observe their interactions closer and have more interactive discussions, which are crucial for refining the game’s design and evaluating its impact. Table 3.1 shows detailed information about the participants, who have all been anonymized. The sample consisting solely of computer science stu-

dents provides benefits while introducing certain constraints. The participants' technical knowledge may lead to different trust levels or verification strategies compared to those with less AI familiarity. They may also be more skeptical of AI-generated content due to their understanding of how machine-learning models work, therefore, their concerns and critiques may not reflect those of the general public. However, this also means that they can provide more informed critiques of AI-generated responses, making their feedback valuable for refining the game's effectiveness in teaching critical evaluation. The participants' familiarity with technology means they would also be able to identify flaws in the game mechanisms and design that may not be as apparent to other users.

Participant Number	Age	Gender	Degree
P1	23	Male	Computer Science
P2	23	Male	Computer Science
P3	21	Male	Computer Science and AI
P4	21	Female	Computer Science and AI

Table 3.1: Research participant summary information

P1 has not taken any AI or Machine Learning (ML) courses. He tends to trust AI chatbots with personal information. He uses a multitude of AI chatbots for different use cases. For example, he uses Claude when providing more sensitive information to the AI chatbot due to the company's strong stance on data privacy and protection. He also uses Cursor for programming support and Perplexity to see the "thought process" behind the generated response. For him, trust in an AI chatbot means trusting that whatever he provides will not be used to train the model further and that the responses provided will be correct and accurate. He is also completing a Large Language Model (LLM) related dissertation.

P2 has taken one AI course. For him, trust in AI tools means trusting that it will generate correct and true information. As he has seen many AI hallucinations, he has a general mistrust of AI and ML systems but still uses them for random tasks and programming support. He mainly uses Claude due to its reputation for being the "best" LLM chatbot.

P3 has taken AI and ML courses and thus understands how ML models work. To him, trust in AI chatbots means trusting that the information generated is true and accurate. As he has seen AI hallucinations, he has a general distrust of ML and AI systems. He still uses AI chatbots for tasks that save him time, such as summarizing documents and providing templates for emails and code. He mainly uses Claude but also DeepSeek for questions relating to AI and ML, as he has found that DeepSeek is good at responding to questions in that domain. He used to use ChatGPT as well.

P4 has taken AI and ML courses and thus understands how ML models work. For her, trust in AI chatbots means trusting that the information provided will be true and accurate and that your personal information will not be used to train the models further. She uses AI chatbots for studying and programming support and smaller tasks such as improving her writing or writing emails. She does not trust AI chatbots with her personal information and thus uses Claude for its strong stance on data protection and privacy. She used to use ChatGPT as well.

3.3 Session 1: GenAI Usage and Prototype Evaluation

3.3.1 Semi-Structured Focus Group

I employed semi-structured group interviews to explore participants' interactions with Generative AI (GenAI) chatbots, particularly focusing on nutrition-related contexts. The interview lasted approximately forty-five minutes and was audio-recorded for subsequent analysis.

I followed a three-part structure that progressively narrowed in focus. We began by exploring participants' general usage patterns of GenAI chatbots, including frequency of use, typical use cases, and methods employed to verify and assess the trustworthiness of AI-generated content. This provided valuable context regarding participants' overall relationship with AI systems. The second phase of the interviews shifted toward the participants' engagement with nutritional information online. Discussions covered the type of nutritional information participants have seen online and their trust in this information. The final section explored the specific intersection of GenAI and nutrition, examining how frequently participants consulted chatbots for nutrition advice and their level of trust in AI-generated nutritional guidance. This gradual approach allowed the participants to reflect on their broader use of AI chatbots before addressing the more specific interactions with nutritional information.

A focus group semi-structured interview format fostered discussions among the participants, allowing them to reflect on each other's perspectives. This method was particularly useful in understanding attitudes towards AI-generated content and potential areas of concern in nutrition-related queries.

3.3.2 Prototype Evaluation and Discussion

The original game concept was straightforward: users create a persona with characteristics that would influence their nutrition, such as dietary preferences or allergens. Players then select query types to receive two answers: one generated by ChatGPT and another that's literature-reviewed. The challenge is to identify the "correct" response. Players would continue to play as long as they wish. I developed a Figma prototype based on this user flow. The query and responses were shown in a phone-like interface. Once the response was selected, a pop-up with an explanation for the response was shown. After exiting the pop-up, the two responses were marked by different colors - red and green. The evaluation of this prototype helped confirm whether the users understand the idea of creating a persona and distinguishing between AI-generated and literature-reviewed responses. The prototype also allowed me to gather feedback on navigation, information presentation, and engagement. I kept the design of the prototype very plain to allow the participants to focus on the mechanics of the game itself rather than the look of the interface. Images of the prototype design can be found in Appendix B.

I employed a think-aloud method in a group setting to allow users to evaluate the Figma prototype of the game. I presented the prototype on a big screen in front of all the participants and asked them to create a persona that was a woman allergic to peanuts and dairy and then one that was a man allergic to wheat. I then asked them to complete one

level in the game. The participants verbalized their thoughts while directing me on how to interact with the system. This facilitated real-time discussions of the game design and mechanics among the participants, providing me with richer insights as users built on each other's observations. This approach also allowed collective problem-solving, as the participants were able to critique and refine ideas together. Observing how the participants negotiated decisions and scrutinized the prototype collectively helped me understand their thoughts about the prototype better. The think-aloud method was followed by a group discussion where I focused on gathering the participants' input on potential gamification elements, features that could enhance the experience, and additional mechanics they would like to see in the system. The session was audio-recorded.

The think-aloud study, followed by a group discussion, provided me with direct insights into participants' cognitive processes while using the system and helped identify usability issues and areas where users may struggle, informing design improvements.

3.4 Session 2: Game Evaluation

3.4.1 Think-Aloud Study

Based on participants' experiences with AI and online nutritional information, along with their evaluation of the initial prototype, I developed a functioning web application. While fully implemented, this system remains a prototype, as it supports gameplay with only a specific set of personas rather than all possible combinations. This limitation is practical, considering that even with restricted persona characteristics, 192 unique personas can be created. As I was writing the literature-reviewed responses, if there were 10 questions per persona, that would mean generating 1920 AI responses, writing 1920 literature-reviewed responses, and 1920 feedback pop explanations, which would have been very time-consuming.

To evaluate the functional prototype, I conducted individual think-aloud sessions with each participant. Participants were asked to complete three key tasks: create a persona, play the game, and check the leaderboard. Throughout the session, participants verbalized their thoughts regarding game mechanics and design elements while navigating the interface. I used screen recordings to capture user interactions and audio recordings, but no recordings of the participants themselves were made.

This was a critical step in assessing the game's perceived educational effectiveness, usability, and overall engagement. The think-aloud method provided me with valuable insights into how users interpreted interface cues, their aesthetic preferences, and their gameplay strategies. These observations directly informed subsequent refinements to the system, enhancing both the educational value and user experience of the game.

3.4.2 Group Discussion

I followed up with a semi-structured group discussion that explored several key areas: the participants' overall user experience with the prototype, navigation, and usability

challenges, the game's perceived effectiveness as an educational tool, and how the experience influenced their trust and understanding of GenAI in nutrition contexts. This format encouraged participants to compare their varying experiences and collectively brainstorm potential improvements. The combination of individual interactions followed by group discussion provided a comprehensive analytical framework, capturing both personal insights and collaborative reflections on the prototype. The session was audio-recorded.

3.5 Data Analysis

3.5.1 Thematic Analysis

I conducted thematic analysis [7] on the transcriptions from both focus group discussions and prototype evaluation sessions. This process began with verbatim transcription of all audio recordings to ensure data integrity. I then read over the transcripts multiple times while noting down any initial observations. For the coding phase, I employed both descriptive and in vivo techniques. Descriptive codes summarize content conceptually, whereas in vivo codes quote the participants' exact words. After generating comprehensive codes, I categorized related codes into potential themes and sub-themes through an iterative process. I created mind maps using the "Miro" software to explore potential relationships between concepts. I then established clear themes by assigning them a name and a detailed description. I selected thematic analysis for its flexibility and established effectiveness in informing meaningful patterns within qualitative research. This approach enabled me to develop robust themes that captured participants' perspectives and experiences. The mind maps of the thematic analysis can be found in Appendix E.

3.5.2 Pain Point Identification

For the functional prototype evaluation sessions, I conducted user journey mapping to document critical incidents and identify pain points. I captured key events during gameplay and user responses by observing users interact with the game interface. During each think-aloud session, I screen-recorded the user interactions and kept note of any facial expressions or body language used to create a comprehensive view of the users' reactions to the interface. I summarized any errors, unexpected occurrences, or misunderstandings into a table in which I categorized usability issues based on frequency (how often issues occurred across multiple users), impact (severity of disruption to gameplay flow and learning objectives), and issue category (whether it's related to UI/UX design or if it is the actual content of the game itself), using a 5-point severity rating scale. This approach led to important discoveries about issues the users faced while interacting with the game. Insights from this analysis informed subsequent design improvements and game mechanics refinements. The table summarizing the pain points can be found in Appendix E.3.

Chapter 4

Focus Group 1: GenAI Usage and Prototype Evaluation

The initial focus group started with a semi-structured group interview followed by prototype testing. During the interview, participants shared their experiences with online nutrition resources and Generative AI (GenAI) tools, particularly focusing on the intersection of these domains. Participants then collaboratively tested a basic prototype, completing tasks together before reflecting on potential improvements to enhance gamification and increase engagement. This dual approach revealed valuable insights into how users interact with and trust GenAI chatbots and online nutrition advice, directly informing the game's design. The prototype testing phase was especially crucial, as it helped understand users' expectations for an educational game and guided the development of core mechanics and features.

4.1 Discussions of Nutrition Misinformation

Participants have noticed that social media platforms have been polluted by nutritional misinformation, unverified claims, and contradictory advice, leading them to express frustration with this confusing landscape of health information online. “Overwhelming,” says P2, capturing the group’s collective frustration with online nutrition content. P2 describes encountering a flood of advice, most lacking scientific support, while P3 cites unsubstantiated claims about nutrition from high-profile political figures, including President Trump’s cabinet members. The discussion reveals a pattern of concern about social media’s role in the spread of nutritional misinformation. P3 observes the periodic cycles of diet trends on social networks such as TikTok that are void of any scientific merit. “On the misinformation front, the diets - [there] are also just so many things that people assume are facts just aren’t really supported by science. [...] What’s led to that being spread so much? Like about your metabolism and how you can get a faster metabolism, et cetera. Like the science behind it just does not back any of these trends that actually just pop up on very prominent social media almost on a periodic yearly basis at this point.” (P3). P4 reinforces this point, describing how influencers sow confusion by frequently contradicting each other, villainizing or

glorifying different foods, supplements, and diets. “I think what’s interesting about TikTok as well because I got so many [videos]. It would be like one guy [who] would be like: “Oh, you should stop eating fruit - that’s bad for you.” And then the next guy would say - “You need to eat a [ton] of fruit” [...]. They’re just fully contradicting each other”. A shared distrust of online nutritional information emerges as participants recount their encounters with unreliable, contradictory, or outright false nutritional information. Many express frustration over the difficulty of finding trustworthy sources. While P1 expresses that they would never trust fitness influencers, preferring guidance from nutrition and exercise science professionals with formal education and certifications, P3 advocates for skepticism towards all nutritional information online. These testimonials highlight a growing wariness of nutritional information online and emphasize the crucial need for critical thinking and reliable sources in an era of misinformation.

4.2 Use of Generative AI Tools

Users’ interactions with AI chatbots can be defined by compromise - a willingness to use them despite recognizing a lack of trust in the technology and a need for verification. The participants displayed diverse approaches to using and sharing information with the AI chatbots. Participants reported extensive use of GenAI chatbots such as ChatGPT or Claude, with P1 candidly admitting to using them “out of laziness” for repetitive tasks like email writing and programming. The applications varied widely, from academic support and coursework implementation to even “medical things.” When discussing the information participants feel comfortable providing to GenAI chatbots, the approaches diverged significantly. Some maintained strict boundaries, providing only basic, non-personal information, while others took a more “liberal” approach, freely sharing documents and even medical test results. “Pretty liberal that to be fair [with providing information]. I specifically use “Claude” which, supposedly, the company values privacy a little bit more and has an explicit setting that doesn’t use your data to train further stuff on. And I guess I take that for granted and just give it whatever. [...] I feel like there’s so much even like health information I’ve given it, like blood tests and scans, and stuff like because it’s not much” (P1). Despite their regular use of these tools, participants expressed a consistent level of skepticism. While they generally trust the chatbots for basic information and general topics, they still approach some topics with caution, noting that they “don’t trust it [chatbots] off the bat.” This skepticism leads them to verify information, particularly in cases involving mathematical calculations, niche topics, or events after 2010. P3 discusses how GenAI responses typically contain a “core fact” surrounded by additional information, emphasizing the importance of identifying and verifying this central claim. “Assuming it’s answering a question, it’s going to try to answer it with some form of very confident statement or fact. And I think the actual skill with using generative AI is to be able to identify what that statement is and just fact-check that one statement” (P3). Participants described using multiple sources for verification, including Google, academic papers, and news sources, though P4 noted the importance of ensuring the credibility of something like a news source itself. This complex relationship with GenAI chatbots - characterized by frequent use despite persistent skepticism - highlights an understanding of the benefits of these tools while remaining cautious of their limitations.

4.3 Nutrition and Generative AI

Although participants reported engaging with AI for various purposes when it comes to nutritional and health advice, the participants displayed reluctance to trust AI tools, strongly preferring traditional information sources such as research papers. The focus group participants reported limited experience with generative AI chatbots in the nutrition domain. Their exposure consisted of encountering AI-generated recipes on social media and using ChatGPT for basic tasks like finding ingredient substitutions and meal inspiration. However, P3 and P4 highlighted that they never used GenAI chatbots in the field of nutrition. The participants' reluctance to use AI for nutritional advice emerged clearly during the discussion. "I've never [asked a GenAI chatbot about nutrition]. I don't think I will, just because I don't think I trust anything it would say. I'd much rather just research it myself on Google" (P4). P1 expressed strong skepticism about trusting AI chatbots with health and nutrition-related topics, stating that even if the AI provided source citations, they can be fabricated or unreliable references. While P2 considered AI trustworthy for simple tasks like ingredient substitutions, P1 noted that the ability to verify the accuracy of such suggestions made them more acceptable. "I have trusted [it] for stuff like: 'In this recipe, I want to substitute this [item]. What ingredients could I use to do this instead?'" So there are... There are a few things that I would generally be like, yeah, this – (I would) be willing to give it a try" (P2). These perspectives highlight a common thread among the participants: a preference for using AI for low-stakes nutritional topics while remaining cautious about its reliability.

4.4 Prototype Evaluation and Suggestions for Improvement

The following sections detail participants' opinions on the Figma prototype. Detailed images of the prototype can be found in Appendix B.

4.4.1 Persona Creation

In the prototype, users had to create a persona before accessing the game. P1 suggested offering a randomly generated persona option, which would streamline gameplay and allow users to experiment with different personas without investing time in the creation process.

4.4.2 AI-Generated vs Literature-Reviewed Response Highlighting

Participants noted that using green and red colors implies correct or incorrect responses, which could be misleading since AI-generated content isn't inherently good or bad. They suggested replacing this color coding with alternative markers that avoid such implications.

4.4.3 Answer Selection

Participants assumed the “correct response” was the literature-reviewed one. They recommended clearer selection instructions. P3 suggested varying the instructions periodically — sometimes asking users to select the AI-generated response, other times the literature-reviewed response — explaining, “Because how I would approach finding the correct literature-reviewed answer is probably different to how I would find an AI answer.”

4.4.4 Response Interface

The prototype displayed responses vertically in a phone-like chat interface. P1 suggested a side-by-side layout would make comparing responses visually easier. P2 proposed using two adjacent phone interfaces, while P4 recommended a two-column format. P3 suggested adopting a tablet-like interface to accommodate side-by-side columns more effectively.

4.4.5 Explanation Pop-Ups

Participants requested that explanation pop-ups include citation sources used in literature-reviewed responses. An example query in the prototype was asking for a meal plan. P1 suggested adding functionality to save the literature-reviewed meal plan within the game, making it accessible on the website.

4.4.6 Countdown and Points

When asked about gamification, two participants simultaneously suggested a “countdown” feature. Others advocated for a points system, with P4 proposing a combination where faster responses earn more points. However, P1 cautioned that this creates a tension between speed and detail, potentially altering gameplay if users feel rushed. P2 recommended implementing a “streak” feature to track consecutive correct answers.

4.4.7 Number of Questions and End Goal

Participants unanimously agreed that the game needs a clear end goal to maintain player engagement. They recommended implementing a fixed number of questions rather than the prototype’s unlimited format. P1 suggested displaying a progress indicator showing completed and remaining questions to help keep the user engaged.

4.4.8 Question Difficulty

P4 recommended implementing progressive difficulty as players advance through questions. P1 elaborated that this difficulty escalation should occur implicitly in the background rather than using explicit “easy,” “medium,” or “hard” levels.

4.5 Key Takeaways and Implications for Game Design

The focus group findings reveal several interconnected themes regarding users' relationships with online nutrition information and generative AI technologies. The participants' responses highlight a complex landscape where online nutrition content and emerging AI tools intersect, characterized by opportunities and significant challenges. A prominent theme emerging from the focus group was the overwhelming nature of online nutrition information, particularly on social media platforms. Participants expressed frustration with the volume of nutrition content and described the contradictory claims, unsubstantiated advice, and periodic diet trends saturating the social media platforms. The participants highlighted the struggle to find trustworthy sources for nutritional information and expressed a lack of trust in influencers and social media in the nutrition domain, preferring advice from exercise or nutrition professionals. Regarding generative AI, participants demonstrated a sophisticated approach to these tools. While they regularly use AI chatbots for various tasks, their interactions are characterized by skepticism - they employ AI for routine tasks while verifying the responses. This approach extends to nutrition-related applications, where participants showed reluctance to rely on AI for advice. The intersection of these themes, distrust in online nutrition information, and cautious engagement with AI presents both challenges and opportunities for the development of the educational game. Participants' experiences suggest that while there is a clear demand for more reliable nutrition information sources, any AI-powered solution must overcome significant trust barriers. The participants' emphasis on verification and source credibility indicates that the game should prioritize clear attribution of information sources. The responses also suggest that the game must not only provide accurate information but also help users develop critical evaluation skills and effectively showcase the benefits of AI tools as a supplement to, rather than a replacement for, professional expertise. The discussion helps clarify key considerations for the development of the game, while also providing insight into the potential users' previous experiences with nutrition and GenAI tools. Mind maps of the thematic analysis can be found in Appendix E.1.

Chapter 5

Game Design and Development

This chapter describes the development of the functioning prototype used in the second focus group sessions. It details the technology stack I used to build the cohesive full-stack application, along with the implementation of the interface design and game mechanics. The design decisions are supported by Human-Computer Interaction (HCI) principles, specifically Nielsen’s heuristics and Gestalt principles, as well as user feedback. Nielsen’s 10 heuristics are general guidelines or “rules of thumb” used to evaluate and improve digital interface design [34]. Gestalt principles are rules that explain how humans perceive visual elements, which guide the compositions of digital interfaces [17]. These HCI principles help formalize justifications for design decisions made, ensuring the use of best practices for user experiences rather than arbitrary aesthetic preferences. Detailed images of the functional prototype developed following the initial focus group can be found in Appendix C.

5.1 Technology Stack

I used a modern web development stack centered around Next.js with its App Router for both frontend and backend functionality. This enabled me to perform file-based routing and provided a seamless way to build both the client-side UI and the serverless API routes within the same codebase. I used TailwindCSS for styling, ensuring consistent design throughout the system. The backend relies on MongoDB (via Mongoose) for data persistence, which provides a flexible, document-based database that works well with TypeScript objects. This technology stack allowed me to create a full-stack application on a single platform without managing separate frontend and backend codebases.

5.2 Game Theme and Overall Design

The topics of nutrition, health, and AI can often feel intimidating and overwhelming. Many people experience anxiety or shame when confronted with health or nutritional information [21]. I designed the game interface to be inviting and friendly, aiming to reduce the stigma surrounding these subjects. The charming food-based characters I drew and animated throughout the game create an emotional connection through

their expressiveness and humor, establishing a consistent design while creating a cute and approachable interface for users to interact with. This design choice actively counters the perception that conversations about nutrition are shameful by adding a humorous element to it. My goal was to craft a welcoming environment for learning without pressure and to ensure the gameplay feels rewarding rather than obligatory and judgmental.

5.3 Interface Features and Game Mechanics

5.3.1 Before Gameplay

5.3.1.1 Username

To enable a leaderboard, I implemented a username system that randomly generates a username by combining random adjectives with food items (for example, “LazyLime” or “ShadyAvocado”). Players can regenerate a username as many times as they would like. In contrast, returning players are prompted to reuse their previous username, preserving their identity across multiple sessions and leaderboard appearances. This approach enhances security by eliminating user inputs, thus preventing NoSQL injection attacks and other MongoDB-specific vulnerabilities. Since users never directly enter text that reaches the database, I don’t need to implement complex input sanitization or validation that would be otherwise necessary.

5.3.1.2 Persona Creation

To protect user privacy, the user creates a fictional persona. Users create these personas by selecting characteristics that influence nutritional recommendations. Although these characteristics could reflect the players themselves, since no other personal information is collected, the persona cannot be linked to the player. For the functional prototype, I implemented three characteristic categories: exercise level, allergens, and dietary preferences. Exercise levels include three options (low, medium, high), from which users must select one. The allergen category offers five options (peanuts, dairy, eggs, fish, and gluten), allowing users to select zero, one, or two allergens. Dietary preferences include three options (vegetarian, vegan, and pescatarian), with users able to select either none or one preference. The system includes logical constraints, such as preventing users with fish allergies from selecting the pescatarian dietary preference. As a user selects different characteristics of the persona, a profile image is created on the right-hand side. Different exercise levels will bring up images of different animals, for example, an otter for a low exercise level. Different allergens will add accessories to the animals, such as a top hat for an egg allergy or a balloon for a gluten allergy. The dietary preferences add a background, for example, mountains for the vegetarian option. The profile image does not serve any other purpose besides engaging the user and adding entertaining features to the game. I chose these characteristics as they have demonstrated an impact on AI-generated nutritional recommendations, as illustrated by [8] and [40]. The persona creation pages can be seen in Figure 5.1 and Figure 5.2.

5.3.1.3 Toggle Switches

When creating personas, users select features through toggle switches. This design implements Nielsen's fifth heuristic [34] - error prevention by clearly indicating selection status, preventing contradictory choices, and ensuring users deselect options before making new selections. The layout also follows the Gestalt principle of similarity [27], with toggle switches grouped in consistent shape, color, and placement. These visual cues help users recognize related options, reinforcing connections between characteristics and simplifying the selection process. An example of the toggle switches can be seen in Figure 5.1.

5.3.1.4 Instructions

Before gameplay, users are presented with comprehensive instructions detailing both the game mechanics and interface elements. This feature meets Nielsen's tenth heuristic [34] - help and documentation by providing users with clear steps to successfully engage with the game. The instructions are focused on the user's tasks and presented before the play begins, ensuring users can navigate the game effectively without excessive trial and error.

5.3.2 During Gameplay

5.3.2.1 Messaging App Interface

The gameplay interface is similar to that of a messaging application, which meets Nielsen's second heuristic [34] - a match between the system and the real world. The feature leverages a familiar digital interaction pattern that most users would already understand. The chat-like interface utilizes users' existing mental models of how messaging applications function, a format they likely use daily. The interface includes recognizable elements such as message bubbles, profile pictures, and a clear distinction between the query and the responses. The familiarity of the interface allows users to focus on the gameplay rather than struggle with how to interact with the game. The messaging app-like interface can be seen in Figure 5.3.

5.3.2.2 Levels

Rather than unlimited play, I implemented a structured progression of 10 questions with escalating difficulty, as suggested by the participants in the initial focus group. This approach allows users to build confidence with simpler examples before introducing more challenging ones. Four questions were drawn directly from Kirk et al. [26], representing some of the most frequent inquiries reported by dietitians, such as:

- What is a healthy snack?
- Is it better to swap sugar for honey? Or sweeteners?
- Are carbohydrates bad?
- Are the sugars in fruit bad to consume?

Two questions were inspired by Haman et al. [19], designed to test players' ability to identify inconsistencies in AI-generated macronutrient information:

- What are the macronutrients in my lunch which is ... (changed depending on the persona)?
- Can you provide a breakfast option that has more than X grams of protein per portion?

The remaining questions address common nutritional topics found online:

- Can you give me a recipe for lunch?
- How many calories should I be consuming in a day?
- What are the benefits of not eating after 8pm (20:00)?
- Will drinking green tea help me burn fat?

The easier questions require minimal nutrition knowledge, while the more difficult ones demand specific expertise or the ability to identify AI-generated text. This approach makes it more accessible for beginners and ensures that users develop critical thinking in this context in a progressive manner. A set number of questions can also provide a sense of accomplishment for players.

5.3.2.3 Questions and Feedback

For each question in the game, I collected AI-generated responses by querying OpenAI's GPT-4 [36] model using a consistent format: "I am a {exercise level}, {dietary preference} individual with {allergen} allergy. {Query question}. Keep the response concise." I implemented this conciseness request to reduce cognitive load on users, as ChatGPT typically generates lengthy text that would be more difficult for players to analyze. To maintain research integrity, I created a new OpenAI account with no connection to my identity. I disabled both the setting allowing OpenAI to train their models on my interactions and the feature that preserved chat history. I ran each query in a new chat session. For the literature-reviewed responses, I used medical journals, research papers, academic articles, government databases, and official nutritional guidelines. The responses primarily consist of direct quotes from these sources, with minimal adjustments to improve links between sentences taken from different sections of the same source or different sources. The feedback explanations include citations for all sources used in the literature-reviewed responses, ensuring players can read further if they wish to. An example of a feedback pop-up can be seen in Figure 5.4.

5.3.3 After Gameplay

5.3.3.1 Leaderboard

The focus group participants emphasized the importance of having a clear end goal for the game, which led me to implement a scoring system with a leaderboard. Players earn points based on how fast they answer the question, even if the player answers correctly but has run out of time, they receive no points. Incorrect answers receive no points.

Upon completion, players view a leaderboard displaying the top 10 scores. Players who don't rank in the top 10 receive feedback on their position in the overall leaderboard and the points they missed to break into the leaderboard. According to Landers and Landers [29], leaderboards can increase player motivation to continue playing and encourage performance improvement when players perceive their current results as inadequate [4]. The leaderboard can be seen in Figure 5.5.

5.3.4 Throughout Gameplay

5.3.4.1 Breadcrumb Menu

A breadcrumb menu enhances navigation in a linear web application by allowing users to backtrack through their journey. This feature satisfies two of Nielsen's key heuristics [34]: it shows system status (Heuristic 1) by clearly indicating the user's current location within the site, reducing uncertainty, and it supports recognition rather than recall (Heuristic 6) by displaying the navigational path rather than requiring users to remember it. The visual presentation of the menu, particularly the arrows between sections, relates to the Gestalt principle of Common Fate [27]. This principle states that elements moving or functioning in the same direction are perceived as a group. In the breadcrumb menu, the arrangement of links connected by directional arrows creates a visual cue that users understand as a hierarchical path. The breadcrumb menu can be seen in Figure 5.1.

5.3.4.2 Tooltips

Tooltips throughout the game provide clear explanations and instructions. They prevent errors by clarifying functionality and interaction expectations before users make mistakes, directly supporting Nielsen's fifth heuristic [34] of error prevention by reducing misunderstandings. Simultaneously, the tooltips serve as relevant assistance exactly where needed, aligning with Nielsen's tenth heuristic [34] of help and documentation. By integrating guidance into the interface, the tooltips enhance usability and provide help to the user. An example of a tooltip can be seen in Figure 5.1.

5.3.4.3 Error Messages

I have implemented informative error messages throughout the game. For example, if a user selects fish as an allergy for the persona but then tries to select a pescatarian diet, they will see an error message about this contradictory combination and will be prompted to change it. The error messages help users understand what went wrong and why it happened and suggest a solution, thus ensuring a smoother recovery process, which aligns with Nielsen's ninth heuristic [34] - help users recognize, diagnose, and recover from errors. By providing immediate feedback, the error messages keep users informed about the system's state, ensuring the user knows if their action has been processed, thus meeting Nielsen's first heuristic [34] - visibility of system status. An example of an error message can be seen in Figure 5.2.

5.3.4.4 Exit and Undo Buttons

Throughout the system, users can use exit and undo buttons to halt or reverse action sequences. This feature meets Nielsen's third heuristic [34] of user control and freedom by ensuring users can correct mistakes or change courses of action without being trapped in unintended processes. These buttons guarantee that users maintain control of their interactions, enabling them to explore the game without making irreversible errors. An example of an exit/undo button can be seen in Figure 5.1.

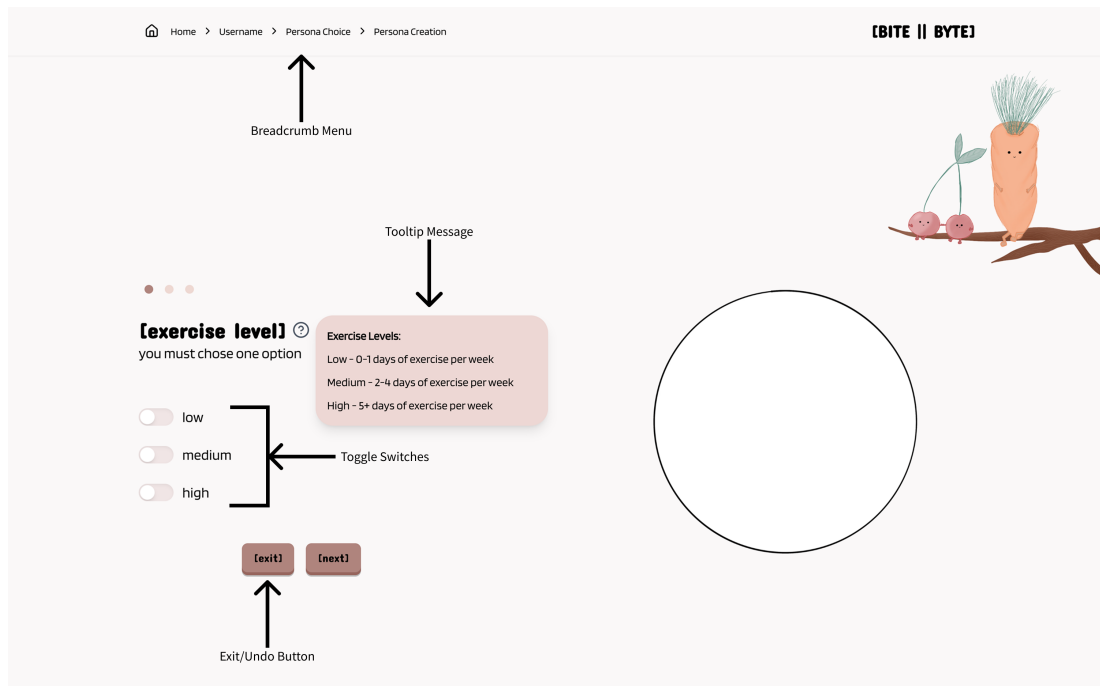


Figure 5.1: The persona creation page at the first step of characteristic selection, highlighting the breadcrumb menu, a tooltip message, toggle switches, and an exit button.

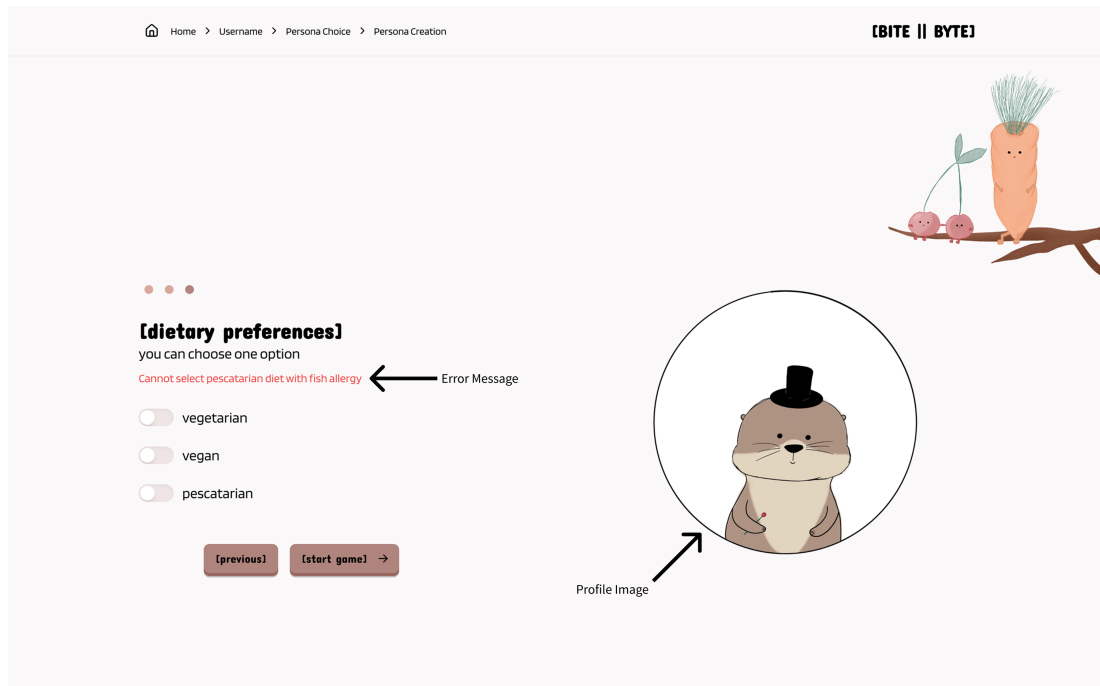


Figure 5.2: The persona creation page at the last stage of characteristic selection, highlighting an error message and the generated profile image thus far.

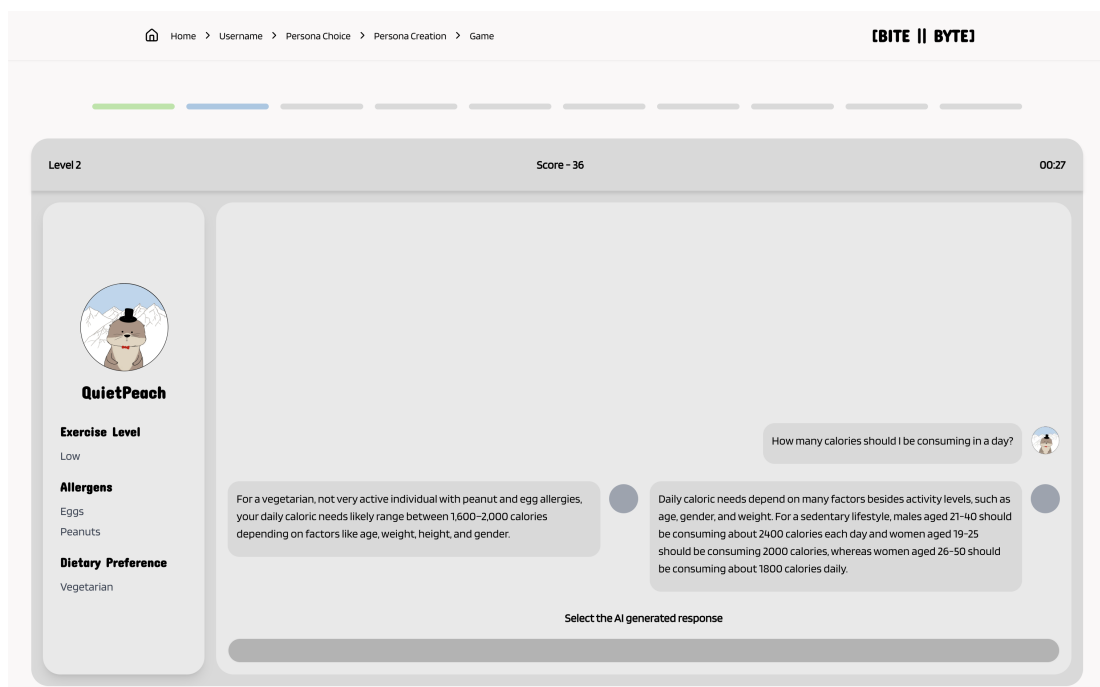


Figure 5.3: The game interface showcasing the messaging app-like interface.

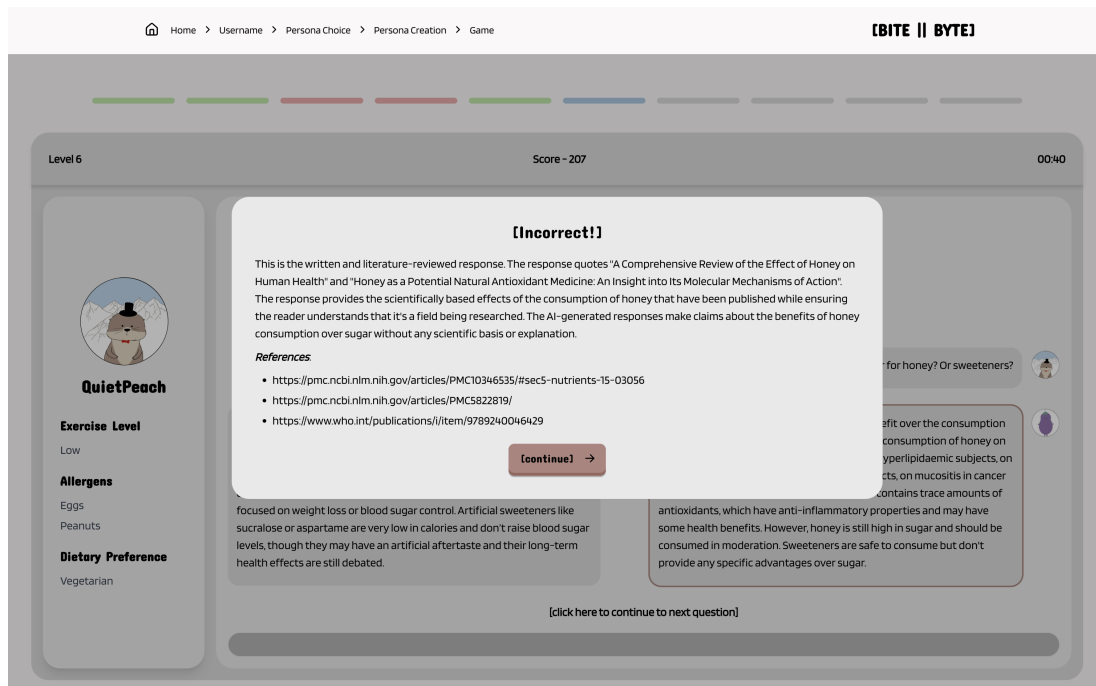


Figure 5.4: The game interface showcasing a feedback pop-up.

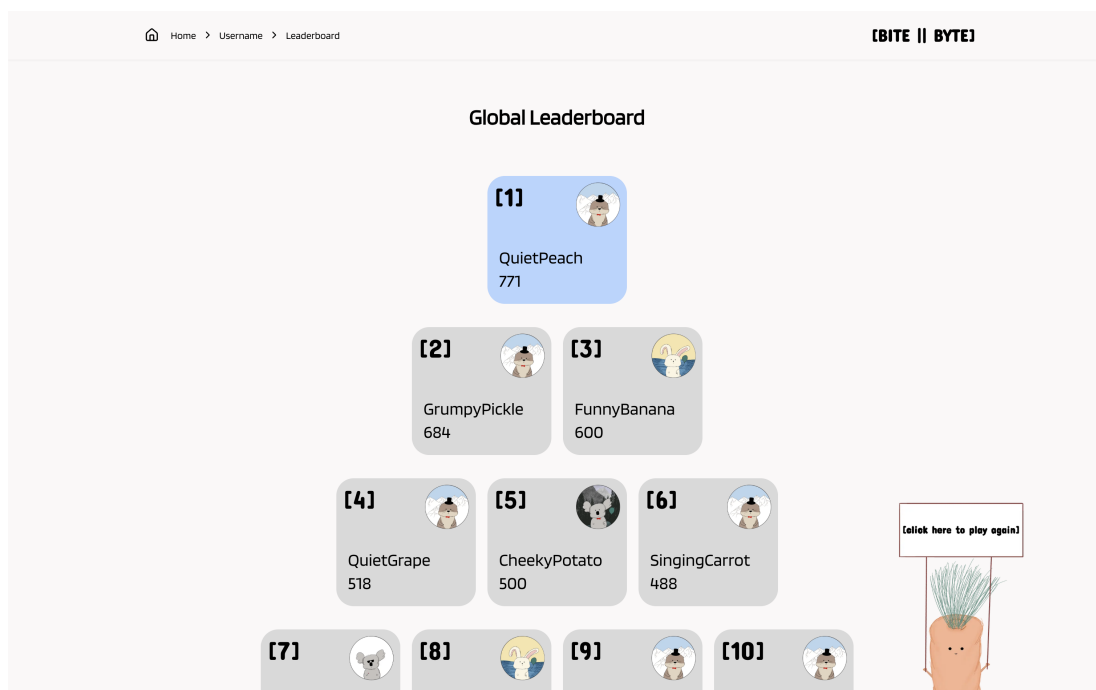


Figure 5.5: The leaderboard page, highlighting the current user who has placed first in the leaderboard.

Chapter 6

Focus Group 2: Functioning Prototype Evaluation

I held think-aloud sessions with each participant individually. During each session, the participants completed the following tasks: create a persona, play the game, and check the leaderboard while expressing their thoughts about the game and the user interface aloud. After the individual sessions, I brought the participants back for a post-play group discussion in which they shared insights about their gameplay experiences, highlighting educational value, interface preferences, and how the game influenced their trust in AI-generated nutritional information. Mind maps of the thematic analysis can be found in Appendix E.2.

6.1 Game Design and Mechanics Feedback

The following sections summarize the feedback received during the individual think-aloud sessions and the post-play group discussion, grouped by different pages of the game.

6.1.1 Landing Page

Only P2 and P3 scrolled through the entire landing page before proceeding to the username page. P2 immediately responded positively to the animations and user interface, exclaiming: “Oh my god, so cute!” while P3 similarly remarked, “Love the landing page.” During the post-task discussion, when asked why they didn’t explore the landing page fully, P1 explained that they hadn’t realized there was more information on the landing page as there was too much white space.

6.1.2 Username Generation

P2 became frustrated when attempting to input a username, not realizing that the username is automatically generated and cannot be directly modified. P1 noted that the explanation text about username functionality was “too blocky” - a problem they found

with most text throughout the game, stating “I’m not going to read all of that.” P3 and P4 did not share any specific reactions to the username generation page.

6.1.3 Persona Creation

Participants provided valuable feedback on the persona creation process. P1 suggested replacing toggles with radio buttons for exercise level selection during persona creation to better indicate that only one option can be selected. P4 found creating their persona intuitive but suggested making certain instructions more explicit (e.g., clarifying the option to select “none” for dietary preferences). Both P4 and P2 expressed amusement regarding the profile picture generation, while P1 appreciated the animation in the top right corner of the page. The participants were instructed to notify me when they felt they had completed each task. All participants except P2 proceeded to the instructions before declaring they had finished, with all three expressing uncertainty about whether the task was actually completed. P1 suggested that a summary of the created persona could be presented before proceeding, which would signal the completion of the persona creation.

6.1.4 Instructions

The animations were appreciated across the group, though P1 found the icon system (pickle for AI-generated and aubergine for literature-reviewed responses) unintuitive. The animated response icons prompted laughter from P3. P1 also found the instructions too dense, recommending to break the text apart more. Despite these concerns, all participants demonstrated clear understanding of the instructions and had no additional significant reactions.

6.1.5 Game

During the post-think-aloud group discussion, all participants expressed appreciation for the messaging app-like interface of the game. P3 specifically highlighted the sidebar, noting that it served as a helpful reminder of their created persona. However, several issues emerged during the think-aloud sessions.

The countdown timer was a significant pain point for all participants. Many ran out of time on questions, with P2 unable to score any points as they ran out of time on all the questions. Most participants admitted that they did not notice the timer or chose to prioritize focusing on the questions instead. As P2 explained, “I’m finding it more interesting to go in-depth than to look at time.” During the group discussion afterward, the participants suggested several solutions: making the timer more visible, extending the time, adding a disappearing border around the messaging interface, or implementing color changes for the final seconds. However, the participants agreed that the best approach would be to replace the countdown with a stopwatch instead. This modification would maintain the time pressure while accommodating players who may require more time for the questions. This feature also raises an interesting question about the balance between the entertainment value and the educational objectives of the game. While a countdown timer is a standard gamification element that adds excitement through time

pressure, in this case, it caused a distraction from the educational aspects. Switching to a stopwatch preserved the gamification element while ensuring the educational objectives of the game were also met.

Throughout the game, the instruction switches from “select the AI-generated response” to “select the non-AI response.” Most participants did not notice this change, making errors when the instructions changed for the first time. After I pointed out the switch, they became more aware of the instructions overall. This feature was polarizing: P1 and P2 felt the switch was unnecessary and added no value to the game, while P3 and P4 argued that it was fine but that the instruction needed to be more visible. Based on this feedback, I decided to eliminate the instructions switching, always asking players to “select the AI-generated response” while making the instruction more prominent and visible to serve as a constant reminder of the task.

After selecting a response, users receive a feedback pop-up that indicated whether their answer was correct. The pop-up also includes guidance on how to discern the AI-generated response and information about any inaccuracies in the AI response. Additionally, it provides an explanation of the literature-reviewed response, and the sources used to write it. Participants responded positively to the feedback pop-ups. P2 found them particularly valuable, stating: “They’re very clear and explain [things] really well.” They suggested making reference links within the pop-ups clickable to create easier access to resources. P3 also emphasized the effectiveness of these pop-ups, noting that after reading them, the correct answer became “extremely obvious.” Throughout their session, P4 adapted their approach to the game based on the feedback pop-ups, highlighting what they were learning. However, they attempted to reopen a closed pop-up, a functionality that I had not implemented. P4 noted that the ability to reopen the feedback pop-up was important and should be incorporated.

All participants except P4 immediately proceeded to the leaderboard after the game and admitted uncertainty about when the game ended. This was a strong indication that the end of the game was not clear.

6.1.6 Leaderboard

P3 stated that the “leaderboard looks great,” while P4 specifically praised the pyramid-like structure, which highlights the top players, as logical. However, P1 expressed confusion about the scoring mechanism, unable to understand what the scores represented or how they were calculated based on factors like time and answer accuracy. Beyond these specific comments, participants did not have specific reactions to the leaderboard.

6.1.7 Other Comments

P3 did not realize that a user may appear in the leaderboard more than once, even with the same username, and suggested clarifying this point in the username generation instructions. P4 expected the game’s logo in the top bar to function as a home button, stating they “assumed that would work.” Participants responded positively to the confirmation alert that appears before exiting the game. P4 enjoyed the game’s

theme, particularly appreciating the drawings and animations, saying that they add “consistency” and describing the design as “cohesive.”

6.2 Post-Play Discussion

During the post-play discussion, participants shared their gameplay experiences and interface preferences, as detailed in the prior sections. They also evaluated the game’s educational value and discussed how it influenced their trust in both AI systems overall and specifically in AI-generated nutritional information and advice. “I did learn a lot,” said P4 when discussing their favorite aspect of the game. P4 also highlighted the profile pictures, describing them as “at the top of the list.” P3 acknowledged the avatars and drawings but agreed with P4, noting they learned a lot from the feedback pop-ups. P2 appreciated the game’s challenging nature. All participants agreed their favorite interface feature was the side-by-side messaging app-like interface of the game. When reflecting on their learning experiences, P3 remarked, “AI is very good at making unsubstantiated claims. Like, I feel that’s something I always knew but I had not seen it in practice as much...” P2 also emphasized how AI can fabricate facts even in very short texts. Since P3 and P4 hadn’t previously used AI for nutritional information, they believe their interactions with AI won’t significantly change. P3 noted that their use cases don’t typically depend on response accuracy, but when they do, they’ll be more wary. P2 stated they would never use generative AI for nutritional help again, saying they’re “absolutely terrified of it.” P2 also expressed having lost considerable trust in generative AI chatbots generally, with P1 agreeing that their approach to AI usage will change due to diminished trust. All participants expressed willingness to play the game again, with P2 adding they would recommend it to acquaintances who use generative AI for nutritional advice without critical evaluation.

6.3 Game Improvements

Images of the game interface changes made after the second focus group can be found in Appendix D. The following were the changes I implemented:

- To create clearer breakpoints between sections of the game, I have implemented confirmation pop-ups that inform the players that they are about to proceed to the next part of the game and request confirmation that the player is ready.
- Players can now reopen feedback pop-ups after closing them, allowing them to review the responses with the information provided in the pop-ups.
- Following industry standards, the game logo in the header now functions as a home button.
- I redesigned the game interface with the messaging app-like interface acting as a fixed frame and internal scrolling, ensuring the interface is accessible on various devices.
- References within the feedback pop-ups are now clickable links that open in new

browser tabs.

- I've improved explanations of the username system and how the score is calculated, as suggested by the participants.
- As participants had differing opinions on the switching between instructions, I decided to remove this functionality. Players are now always asked to select the AI-generated response. Participants also mentioned that the instruction was not noticeable, so I moved it to the top of the interface.
- Instead of the strict countdown, I implemented a stopwatch mechanism instead, which rewards speed without penalizing those that need more time. Players now receive a base of 50 points for each correct answer regardless of timing, with up to an additional 50 points awarded for those who respond within the first minute. This creates a motivating time-pressure and accommodates diverse reading speeds and approaches, resulting in a more inclusive design that maintains competitive elements.

Chapter 7

Discussion

In this study, I explored the intersection between generative AI, nutrition misinformation, and educational game-based learning. The aim was to develop and evaluate an educational game that enhances users' critical literacy skills in distinguishing AI-generated content and promotes safer engagement with GenAI chatbots in nutrition-related contexts. The findings reveal a complex relationship between trust in AI chatbots, skepticism toward online nutritional content, and the potential of game-based learning to enhance AI literacy.

Through thematic analysis, several key themes emerged, including a general lack of trust for AI-generated content with a tendency to verify responses, everyday interactions with nutritional misinformation online, different use cases of AI in nutritional domains, and the role of game mechanics in facilitating learning. During the post-gameplay discussions, participants expressed that their previous ways of discerning AI-generated content had changed and that they had lost trust in AI-generated content, especially in nutrition contexts. These findings contribute to broader discussions on AI's role in nutritional and dietary communication, the challenges of digital and AI literacy, and the value of game-based learning in fostering critical thinking skills.

7.1 GenAI for Nutritional Advice

Despite the frequent use of AI chatbots for various tasks, participants expressed a reluctance to rely on GenAI for nutrition-related queries. Participants discussed a variety of verification strategies when interacting with AI-generated content, such as identifying a “core fact” within the AI response and verifying this information with other sources, such as academic papers. This approach reflects the participants' awareness of possible AI hallucinations, as described by Huang et al. [23]: AI can generate confident-sounding statements that may be factually incorrect. Prior studies have found that AI tools like ChatGPT can provide generally accurate nutritional information, but they can also make detrimental mistakes in more complex cases [35] [40] [19]. By situating an educational game within this domain, this research attempts to bridge the gap between AI's potential and the critical thinking skills necessary for safe and informed usage in nutritional contexts. Kirk et al. [26] investigated “the competency of ChatGPT

in answering common nutrition questions” by having participants compare answers to common nutritional questions answered by both ChatGPT and real nutritionists, where the participants were other dietitians or nutritionists and experts in the domain. The study focused on the correctness, actionability, and comprehensibility of the AI-generated responses [26]. This research expands on the work of Kirk et al. [26], as the game involves a similar mechanic of comparing answers, but with a focus on correctness and the ability to distinguish AI-generated text, aiming to teach people who may not be nutritional experts, about the dangers of incorrect information presented by GenAI. A clear change in trust levels was demonstrated by the participants after the gameplay, with P2 expressing fear of using GenAI for nutritional advice after playing the game, attributing this shift in perspective to seeing the detrimental mistakes sometimes made by GenAI. Although Kirk et al. [26] found that participants found AI-generated responses more actionable and comprehensible, the nature of this study highlights the inaccuracy of GenAI responses to queries in the domain of nutrition. The participants also showcased a growing awareness of subtle hallucination patterns in AI-generated responses. Studies such as Ponzo et al. [40] and Haman et al. [19] found that hallucinations and incorrect information can often be generated by AI chatbots when answering questions including allergens or macronutrients. Participants’ ability to recognize these hallucinations based on the formatting of the response and other clues showcases AI literacy - an emerging form of digital literacy. Tiernan et al. [48] argue that critical evaluation of AI-generated content requires the development of new digital literacy frameworks. The findings highlight the potential of educational games in providing users with ways to learn AI literacy and critical evaluation skills in a safe environment.

7.2 Effectiveness of Game-Based Learning for AI Literacy

The study demonstrated that an interactive, game-based approach can help users develop critical thinking skills for assessing AI-generated nutritional information. Participants found the game engaging and claimed to have learned new skills, particularly from the feedback pop-ups that provided explanations and source citations. This supports existing literature on serious games as effective tools for teaching digital literacy and critical thinking [41]. However, there is still a gap in research on the use of educational games to teach AI literacy, particularly in the context of nutrition. Although AI ethics and security have been explored in some game-based learning approaches, few studies have examined its effectiveness in teaching critical evaluation of AI-generated nutrition content [33] [3]. Recent research does indicate a growing interest in using games to educate users about AI’s limitations, such as Microsoft’s “CyberSade AI: Dig Deeper” with Minecraft [33], Google’s “Odd One Out,” [18], and Ren-AI [3]. These efforts highlight the potential of game-based learning in AI literacy education. By applying similar methodologies to nutrition and dietary advice, an educational game could equip users with skills to recognize AI-generated misinformation, assess AI-generated recommendations critically, and understand the biases of AI models. This study contributes to this growing field as it leverages gamification, incorporating

leaderboards, time-based scoring, and interactive scenarios to engage users, like other games of similar contexts, to foster a more critical approach to AI-generated nutrition content and highlights the need for further research on the effectiveness of such tools.

7.3 Limitations

This study has several limitations that should be considered when interpreting the findings. First, the small sample size of four participants, all with computer science backgrounds, limits the generalizability of the results. The participants' technical knowledge and familiarity with AI systems likely influenced their interactions with the game and their ability to identify AI-generated content in ways that may be different from the general population.

Second, while I carefully selected the nutritional questions to be used within the game, they do not address the full range of nutrition-related topics where users might encounter AI-generated misinformation. The questions are also the same for each persona. Although this makes it easier to compare the experiences of the participants evaluating the game, it lowers the probability that someone will play again.

Third, I only used OpenAI's GPT-4 [36] to retrieve the AI-generated responses for the game. This means the game only represents one GenAI model. Different models may exhibit varying hallucination patterns and accuracy levels for nutritional information. Although this approach doesn't take away from the aim to teach people about the dangers of using GenAI in nutritional contexts, it falls short of fully teaching them to identify AI-generated nutritional content. Different AI models produce distinctive outputs with unique patterns and characteristics, making it more difficult to develop universal identification skills.

Lastly, the written responses in the game were compiled by me, someone without formal education in nutrition. To mitigate potential inaccuracies, I carefully formatted the responses using direct quotes from medical journals and articles, as well as government guidelines. Despite these efforts, there remains a risk of misinterpretation and unintentional bias in the information. Future iteration of the game would benefit from employing expert nutritionists and dietitians to generate written responses, ensuring the game's reliability as an educational tool.

7.4 Further Research

Future research could expand this study in several ways. First, testing the game with a larger and more diverse sample of participants with varying technical backgrounds, nutritional knowledge, and familiarity with AI. This would help produce a more in-depth, generalizable evaluation of the game.

To increase the educational value of the game, it could be expanded to include a wider range of nutritional topics and topics in other areas of healthcare and wellbeing. This could also involve using different AI models to generate the responses, helping users recognize varying patterns of AI-generated content. Additionally, adding more

characteristics to the persona, such as age and medical conditions, could enhance the personalization of the learning experience.

Lastly, dietitians, nutritionists, and medical experts should be employed to generate the written answers to ensure the reliability of the game. Furthermore, teams of nutrition experts and AI system experts should be enlisted to write the explanatory feedback pop-ups for each level of the game, ensuring the users receive accurate explanations that balance nutritional science and AI literacy.

Chapter 8

Conclusions

This study investigated the intersection of GenAI, nutritional misinformation, and game-based learning to develop and evaluate an educational tool that focuses on enhancing users' critical evaluation skills of AI-generated content in nutrition-related contexts.

The initial focus group revealed a complex relationship between users and AI tools. Participants admitted to regularly using AI chatbots for different tasks but maintained a level of skepticism toward the AI-generated responses. The participants recognized the potential for AI hallucinations, as they described using different methods of verification to fact-check AI-generated responses.

An evaluation of an initial prototype guided the development of a functioning game prototype, shaping the design elements and game mechanics. The evaluation of the functioning prototype through individual think-aloud sessions revealed the potential of the education tool in teaching users critical thinking skills and highlighting the dangers of using GenAI in sensitive contexts like nutrition in an engaging way. Key features of the game that contributed to its effectiveness included:

- A messaging app-like interface that provided a familiar experience
- Feedback pop-ups explaining the responses, potential AI hallucinations, inaccuracies, evaluation techniques, and providing references to information used to write the non-AI generated responses.
- Gamification elements such as a leaderboard and scoring system that maintained user engagement.

In a group post-gameplay discussion, participants described a change in their perspective on AI chatbots, including a loss of trust, a more careful approach to the use of AI in health and nutrition domains, and improved skills in recognizing AI-generated content. This suggests the game successfully achieved its objectives.

The study demonstrated the potential dangers of AI hallucinations in nutrition contexts, adding to the research highlighting limitations of existing AI tools in health-related domains and the potential of using game-based approaches to teach AI literacy. The limitations of the study inform potential future work, including expanding the game to

other health-related domains, conducting larger-scale user evaluations, and collaborating with nutrition experts and dietitians to improve the quality of the in-game content.

Bibliography

- [1] Marina Adami. *AI-generated slop is quietly conquering the internet. Is it a threat to journalism or a problem that will fix itself?* 2024. URL: <https://reutersinstitute.politics.ox.ac.uk/news/ai-generated-slop-quietly-conquering-internet-it-threat-journalism-or-problem-will-fix-itself>.
- [2] Anthropic. *Meet Claude*. URL: <https://www.anthropic.com/claude>.
- [3] Mine Arai et al. “REN-A.I.: A Video Game for AI Security Education Leveraging Episodic Memory”. In: *IEEE Access* 12 (2024). DOI: <https://doi.org/10.1109/access.2024.3377699>.
- [4] Gabriel Barata et al. “So Fun It Hurts – Gamifying an Engineering Course”. In: *Lecture Notes in Computer Science* (2013), pp. 639–648. DOI: https://doi.org/10.1007/978-3-642-39454-6_68.
- [5] Katrin Becker. “What’s the difference between gamification, serious games, educational games, and game-based learning?” In: *Academia Letters* (2021). URL: https://www.academia.edu/45044609/What_s_the_difference_between_gamification_serious_games_educational_games_and_game_based_learning.
- [6] Andrew Bouras. *Integrating Randomness in Large Language Models: A Linear Congruential Generator Approach for Generating Clinically Relevant Content*. 2024. URL: <https://arxiv.org/abs/2407.03582v1>.
- [7] Virginia Braun and Victoria Clarke. “Using thematic analysis in psychology”. In: *Qualitative Research in Psychology* 3.2 (2006), pp. 77–101. DOI: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa).
- [8] Angeline Chatelan, Aurélien Clerc, and Pierre-Alexandre Fonta. “ChatGPT and Future Artificial Intelligence Chatbots: What may be the Influence on Credentialed Nutrition and Dietetics Practitioners?” In: *Journal of the Academy of Nutrition and Dietetics* 123.11 (Aug. 2023). DOI: <https://doi.org/10.1016/j.jand.2023.08.001>. URL: <https://www.sciencedirect.com/science/article/pii/S2212267223013084>.
- [9] Mikaël Chelli et al. “Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis”. In: *Journal of Medical Internet Research* 26 (May 2024). DOI: <https://doi.org/10.2196/53164>. URL: <https://pubmed.ncbi.nlm.nih.gov/38776130/>.
- [10] Yung-Sung Chuang et al. *DoLa: Decoding by Contrasting Layers Improves Factuality in Large Language Models*. 2023. URL: <https://arxiv.org/abs/2309.03883>.

- [11] Google Deepmind. *Gemini*. 2024. URL: <https://deepmind.google/technologies/gemini/>.
- [12] DeepSeek. *DeepSeek*. 2025. URL: <https://www.deepseek.com/>.
- [13] Emily Denniss, Rebecca Lindberg, and Sarah A. McNaughton. “Quality and Accuracy of Online Nutrition-Related Information: a Systematic Review of Content Analysis Studies”. In: *Public Health Nutrition* 26.7 (May 2023), pp. 1–13. DOI: <https://doi.org/10.1017/S1368980023000873>. URL: <https://www.cambridge.org/core/journals/public-health-nutrition/article/quality-and-accuracy-of-online-nutritionrelated-information-a-systematic-review-of-content-analysis-studies/4E28F7A056AA8CB4A19F9E5DF0B095C>
- [14] Minecraft Education. *Cybersafe AI Dig Deeper*. 2025. URL: <https://education.minecraft.net/en-us/blog/cybersafe-ai-dig-deeper>.
- [15] Craig E. Emes. “Is Mr Pac Man Eating Our Children? A Review of the Effect of Video Games on Children”. In: *The Canadian Journal of Psychiatry* 42.4 (1997), pp. 409–414. DOI: 10.1177/070674379704200408.
- [16] Europol. *Facing Reality? Law Enforcement and the Challenge of Deepfakes*. Observatory Report. Luxembourg: Europol Innovation Lab, 2022.
- [17] Interaction Design Foundation. *What Are Gestalt Principles?* Aug. 2016. URL: <https://www.interaction-design.org/literature/topics/gestalt-principles>.
- [18] Experiments with Google. *Odd One Out by Caroline Buttet and Emmanuel Durgoni, Artists in Residence at Google Arts Culture Lab - Experiments with Google*. URL: <https://experiments.withgoogle.com/odd-one-out>.
- [19] Michael Haman, Milan Školník, and Michal Lošťák. “AI Dietitian: Unveiling the Accuracy of ChatGPT’s Nutritional Estimations”. In: *Nutrition* 119 (Mar. 2024). DOI: <https://doi.org/10.1016/j.nut.2023.112325>. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0899900723003532>.
- [20] Juho Hamari et al. “Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning”. In: *Computers in Human Behavior* 54 (Jan. 2016), pp. 170–179. DOI: <https://doi.org/10.1016/j.chb.2015.07.045>.
- [21] Alison Hann, Ashley Frawley, and Gillian Spedding. “Not very NICE: deviance, stigma and nutritional guidelines related to healthy weight and obesity”. In: *The International Journal of Health Planning and Management* 32.4 (Apr. 2016), pp. 416–432. DOI: <https://doi.org/10.1002/hpm.2350>.
- [22] Krystal Hu. “ChatGPT sets record for fastest-growing user base”. In: *Reuters* (Feb. 2023). URL: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>.
- [23] Lei Huang et al. “A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions”. In: *ACM transactions on office information systems* 43.2 (Nov. 2024). DOI: <https://doi.org/10.1145/3703155>.
- [24] Dirk Ifenthaler, Deniz Eseryel, and Xun Ge. “Assessment for Game-Based Learning”. In: *Assessment in Game-Based Learning* (2012), pp. 1–8. DOI: https://doi.org/10.1007/978-1-4614-3546-4_1.

- [25] Kristian Kiili. “Digital game-based learning: Towards an experiential gaming model”. In: *The Internet and Higher Education* 8.1 (Jan. 2005), pp. 13–24. DOI: <https://doi.org/10.1016/j.iheduc.2004.12.001>. URL: http://www.savie.ca/sage/articles/940_300027-kiili-2005.pdf.
- [26] Daniel Kirk, Elise van Eijnatten, and Guido Camps. “Comparison of Answers between ChatGPT and Human Dieticians to Common Nutrition Questions”. In: *Journal of Nutrition and Metabolism* 2023 (Nov. 2023), pp. 1–9. DOI: <https://doi.org/10.1155/2023/5548684>.
- [27] Kurt Koffka. *Principles of Gestalt Psychology*. Harcourt, Brace, 1935.
- [28] Fedwa Laamarti, Mohamad Eid, and Abdulmotaleb El Saddik. “An Overview of Serious Games”. In: *International Journal of Computer Games Technology* (2014), pp. 1–15. DOI: <https://doi.org/10.1155/2014/358152>. URL: <https://www.hindawi.com/journals/ijcgt/2014/358152/#B69>.
- [29] Richard N. Landers and Amy K. Landers. “An Empirical Test of the Theory of Gamified Learning”. In: *Simulation Gaming* 45.6 (Dec. 2014), pp. 769–785. DOI: <https://doi.org/10.1177/1046878114563662>.
- [30] Potsawee Manakul, Adian Liusie, and Mark J. F. Gales. “SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models”. In: *arXiv:2303.08896* (Mar. 2023). URL: <https://arxiv.org/abs/2303.08896>.
- [31] Amarachi B. Mbakwe et al. “ChatGPT passing USMLE Shines a Spotlight on the flaws of Medical Education”. In: *PLOS Digital Health* 2.2 (Feb. 2023). Ed. by Harry Hochheiser. DOI: <https://doi.org/10.1371/journal.pdig.0000205>.
- [32] Microsoft. *Microsoft Copilot for Microsoft 365 Overview*. Mar. 2024. URL: <https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-overview>.
- [33] Microsoft. *Safer Internet Day 2025: Tackling abusive AI-generated content risks through education and empowerment – Singapore News Center*. Feb. 2025. URL: <https://news.microsoft.com/en-sg/2025/02/18/safer-internet-day-2025-tackling-abusive-ai-generated-content-risks-through-education-and-empowerment/>.
- [34] Jakob Nielsen. “Enhancing the Explanatory Power of Usability Heuristics”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94* (1994), pp. 152–158. DOI: <https://doi.org/10.1145/191666.191729>. URL: <https://dl.acm.org/citation.cfm?id=191729>.
- [35] Paweł Niszczoła and Iga Rybicka. “The Credibility of Dietary Advice Formulated by ChatGPT: Robo-diets for People With Food Allergies”. In: *Nutrition* 112 (May 2023). DOI: <https://doi.org/10.1016/j.nut.2023.112076>.
- [36] OpenAI. “GPT-4 Technical Report”. In: *arXiv:2303.08774* (Mar. 2023). DOI: <https://doi.org/10.48550/arXiv.2303.08774>. URL: <https://arxiv.org/abs/2303.08774>.
- [37] OpenAI. *Introducing ChatGPT*. Nov. 2022. URL: <https://openai.com/index/chatgpt/>.

- [38] World Health Organization. *Noncommunicable Diseases*. Dec. 2024. URL: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>.
- [39] Jan L. Plass, Bruce D. Homer, and Charles K. Kinzer. “Foundations of game-based learning”. In: *Educational Psychologist* 50.4 (2015), pp. 258–283. DOI: <https://doi.org/10.1080/00461520.2015.1122533>. URL: <https://files.eric.ed.gov/fulltext/EJ1090277.pdf>.
- [40] Valentina Ponzo et al. “Is ChatGPT an Effective Tool for Providing Dietary Advice?” In: *Nutrients* 16.4 (Jan. 2024). DOI: <https://doi.org/10.3390/nut16040469>. URL: <https://www.mdpi.com/2072-6643/16/4/469>.
- [41] Meihua Qian and Karen R. Clark. “Game-based Learning and 21st century skills: A review of recent research”. In: *Computers in Human Behavior* 63.63 (Oct. 2016), pp. 50–58. DOI: <https://doi.org/10.1016/j.chb.2016.05.023>.
- [42] Claudia Schrader. “Serious Games and Game-Based Learning”. In: *Handbook of Open, Distance and Digital Education* (2022), pp. 1–14. DOI: https://doi.org/10.1007/978-981-19-0351-9_74-1.
- [43] David Williamson Shaffer. *How Computer Games Help Children Learn*. New York: Palgrave Macmillan US, 2006. ISBN: 9780230602526. DOI: <https://doi.org/10.1057/9780230601994>.
- [44] Ilia Shumailov et al. “AI models collapse when trained on recursively generated data”. In: *Nature* 631.8022 (July 2024), pp. 755–759. DOI: <https://doi.org/10.1038/s41586-024-07566-y>.
- [45] Ilia Shumailov et al. *The Curse of Recursion: Training on Generated Data Makes Models Forget*. May 2023. DOI: <https://doi.org/10.48550/arXiv.2305.17493>. URL: <https://arxiv.org/abs/2305.17493>.
- [46] Tarja Susi, Mikael Johannesson, and Per Backlund. *Serious Games: An Overview*. 2024. URL: <https://www.diva-portal.org/smash/record.jsf?pid=diva2:2416&dswid=-3373>.
- [47] Brian Thompson et al. *A Shocking Amount of the Web is Machine Translated: Insights from Multi-Way Parallelism*. June 2024. URL: <https://arxiv.org/pdf/2401.05749>.
- [48] Peter Tiernan et al. “Information and Media Literacy in the Age of AI: Options for the Future”. In: *Education Sciences* 13.9 (Sept. 2023). DOI: <https://doi.org/10.3390/educsci13090906>. URL: <https://www.mdpi.com/2227-7102/13/9/906>.
- [49] Vasileios Tsiantis, Dimitrios Konstantinidis, and Kosmas Dimitropoulos. “Chat-GPT in Nutrition: Trends Challenges and Future Directions”. In: (June 2024). DOI: <https://doi.org/10.1145/3652037.3663898>.
- [50] Sowmya Vajjala et al. *Practical Natural Language Processing : A Comprehensive Guide To Building Real-World NLP Sysems*. O’Reilly Media, Inc., 2020. ISBN: 9781492054054.
- [51] Jim Waldo and Soline Boussard. “GPTs and Hallucination”. In: *Queue* 22.4 (Aug. 2024), pp. 19–33. DOI: <https://doi.org/10.1145/3688007>.

Appendix A

Supporting Ethics Documents

A.1 Participant Information Sheet

The participant information sheet can be seen in Figures A.1 A.2 A.3 A.4

A.2 Participant Consent Form

The participant consent form can be seen in Figure A.5

Participant Information Sheet

Project title:	Topics In Digital Wellbeing
Principal investigator:	Tara Capel
Researcher collecting data:	Roberta Posiunaite
Funder (if applicable):	

This study was certified according to the Informatics Research Ethics Process, reference number 774516. Please take time to read the following information carefully. You should keep this page for your records.

Who are the researchers?

Roberta Posiunaite is a 4th year Computer Science student.

Tara Capel is the project supervisor.

What is the purpose of the study?

This research aims to understand how users interact with generative AI (GenAI) tools such as ChatGPT in the context of nutrition. It seeks to answer questions about what questions users ask, how users feel after receiving responses, and whether users trust the responses. The focus is on understanding the users' interactions and feelings and evaluating a prototype of an educational tool built around the use of GenAI in the sphere of nutrition. The educational tool is a game where users create a persona and then choose between different queries. The user will receive two outputs, one pre-generated by ChatGPT, and another one verified by literature. The user will have to find the ChatGPT-generated one. The system itself will never interact with ChatGPT, all the responses will be pre-generated based on the different combinations of persona characteristics and different types of queries.

Why have I been asked to take part?

You have been asked to take part in this research as someone who is interested in or uses GenAI tools such as ChatGPT. You may also be someone interested in nutrition or use GenAI tools in the context of nutrition.

Do I have to take part?

Figure A.1: The Participant Information Sheet Page 1

No – participation in this study is entirely up to you. You can withdraw from the study at any time, up until December without giving a reason. After this point, personal data will be deleted and anonymised data will be combined such that it is impossible to remove individual information from the analysis. Your rights will not be affected. If you wish to withdraw, contact the PI. We will keep copies of your original consent, and of your withdrawal request.

What will happen if I decide to take part?

The research will be conducted in two sessions. The first session will be an in-person workshop-style discussion in December. You will be asked some questions about your interactions with GenAI tools such as ChatGPT, with a focus on your interactions involving nutrition. You will then be asked to interact with a prototype of a system, where you will be asked to talk through your thoughts and feelings about it. The second session will be an in-person interview in January. You will be asked to evaluate a more complete version of the system, by talking through your interactions with the system. The system is an educational game, where you will create a persona based on different characteristics. You will then be able to formulate a query and will be presented with two answers, one generated by ChatGPT, and one verified by literature. Your task will be to find the ChatGPT-generated answer. Both sessions will not be longer than three hours each. Audio of the sessions will be recorded. Video recordings will not be taken.

Are there any risks associated with taking part?

There are no significant risks associated with participation.

Are there any benefits associated with taking part?

No.

What will happen to the results of this study?

The results of this study may be summarised in published articles, reports and presentations. Quotes or key findings will be anonymized: We will remove any information that could, in our assessment, allow anyone to identify you. With your consent, information can also be used for future research. Your data may be



Figure A.2: The Participant Information Sheet Page 2

archived for a maximum of four years. All potentially identifiable data will be deleted within this timeframe if it has not already been deleted as part of anonymization.

Data protection and confidentiality.

Your data will be processed in accordance with Data Protection Law. All information collected about you will be kept strictly confidential. Your data will be referred to by a unique participant number rather than by name. Your data will only be viewed by the researcher/research team – Roberta Posiunaite and Tara Capel.

All electronic data will be stored on a password-protected encrypted computer, on the School of Informatics' secure file servers, or on the University's secure encrypted cloud storage services (DataShare, ownCloud, or Sharepoint) and all paper records will be stored in a locked filing cabinet in the PI's office. Your consent information will be kept separately from your responses in order to minimise risk.

What are my data protection rights?

The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

Who can I contact?

If you have any further questions about the study, please contact the lead researcher, Roberta Posiunaite s2158723@ed.ac.uk.

If you wish to make a complaint about the study, please contact inf-ethics@inf.ed.ac.uk. When you contact us, please provide the study title and detail the nature of your complaint.

Updated information.

If the research project changes in any way, an updated Participant Information Sheet will be made available on <http://web.inf.ed.ac.uk/infweb/research/study-updates>.

Alternative formats.

Figure A.3: The Participant Information Sheet Page 3

Page 4 of 4

To request this document in an alternative format, such as large print or on coloured paper, please contact Roberta Posiunaite s2158723@ed.ac.uk.

General information.

For general information about how we use your data, go to: edin.ac/privacy-research



Figure A.4: The Participant Information Sheet Page 4

Participant number: _____

Participant Consent Form

Project title:	Topics in Digital Wellbeing
Principal investigator (PI):	Tara Capel
Researcher:	Roberta Posiunaite
PI contact details:	tcapel@ed.ac.uk

By participating in the study you agree that:

- I have read and understood the Participant Information Sheet for the above study, that I have had the opportunity to ask questions, and that any questions I had were answered to my satisfaction.
- My participation is voluntary, and that I can withdraw at any time without giving a reason. Withdrawing will not affect any of my rights.
- I consent to my anonymised data being used in academic publications and presentations.
- I understand that my anonymised data will be stored for the duration outlined in the Participant Information Sheet.

Please tick yes or no for each of these statements.

1. I agree to being audio recorded.

YesNo

2. I allow my data to be used in future ethically approved research.

YesNo

3. I agree to take part in this study.

YesNo

Name of person giving consent

Date

Signature

dd/mm/yy

Name of person taking consent

Date

Signature

dd/mm/yy



Figure A.5: The Participant Consent Form

Appendix B

Initial Figma Prototype

The following images detail the Figma prototype presented to the participants during the initial focus group.

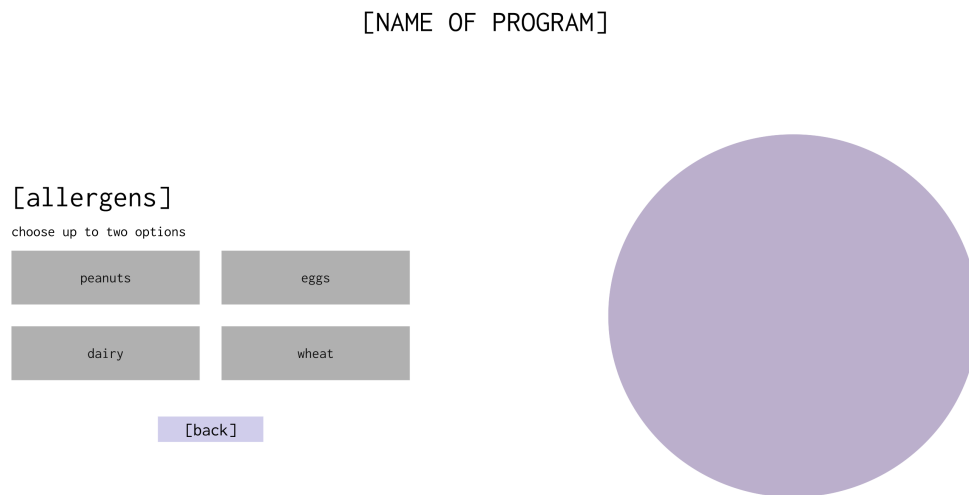


Figure B.1: The first page participants were presented with the Figma prototype.

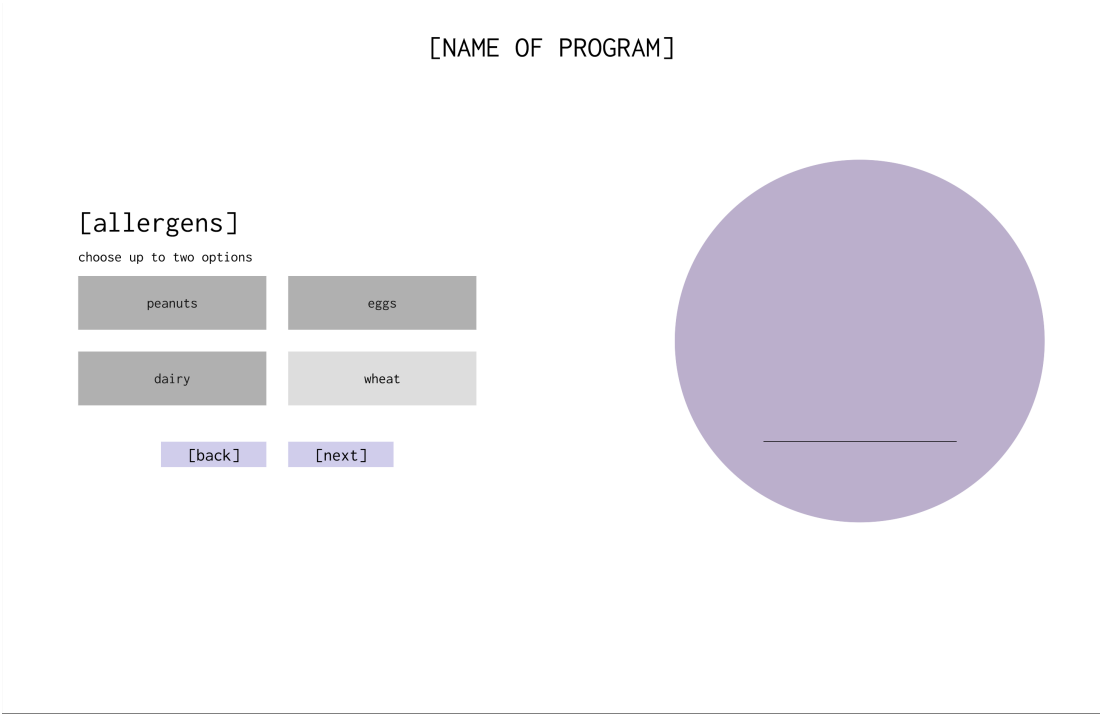


Figure B.2: The first page of the prototype after “wheat” is selected.

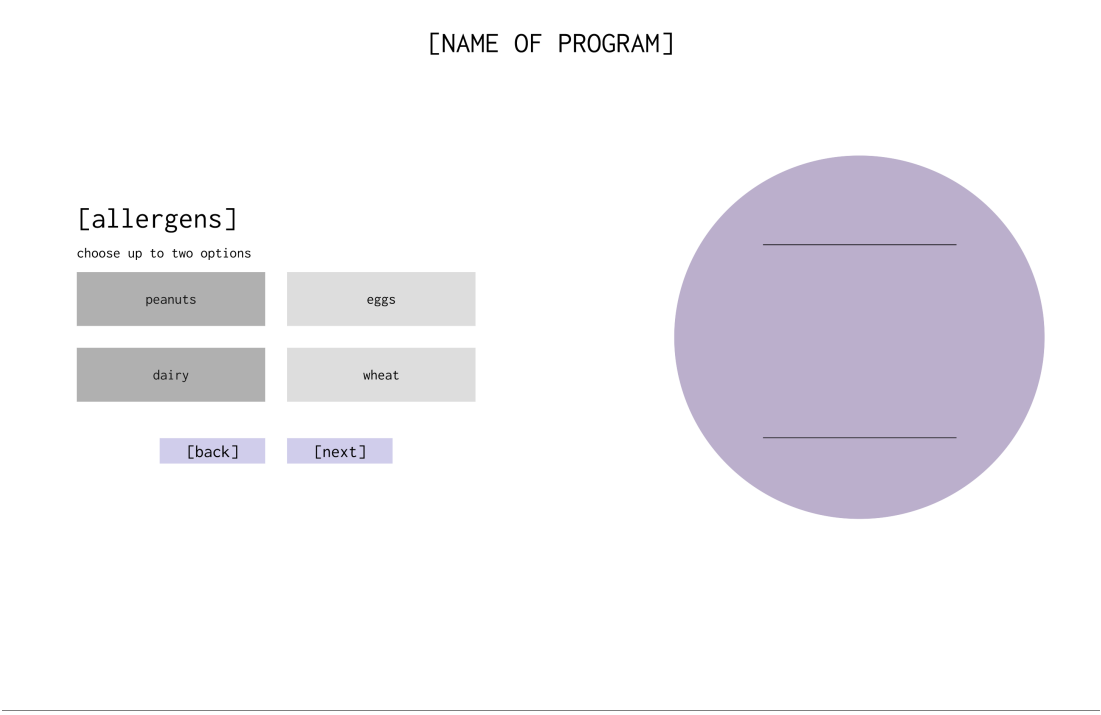


Figure B.3: The first page of the prototype after “wheat” and “eggs” have been selected.

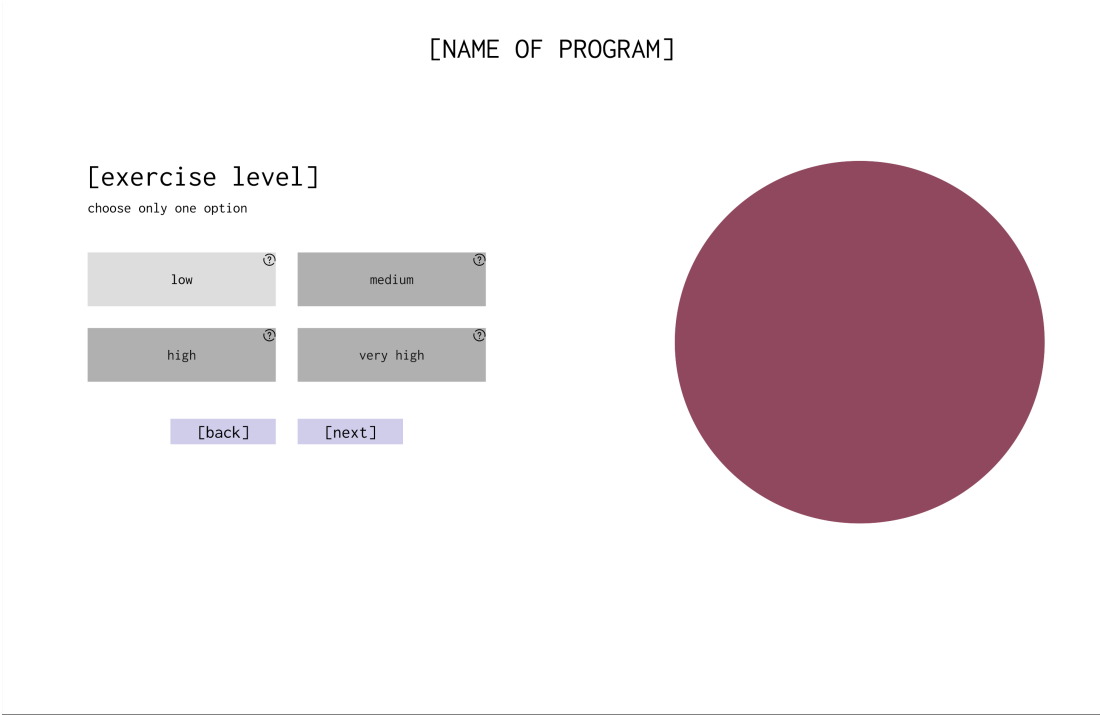


Figure B.4: The second page of the Figma prototype.

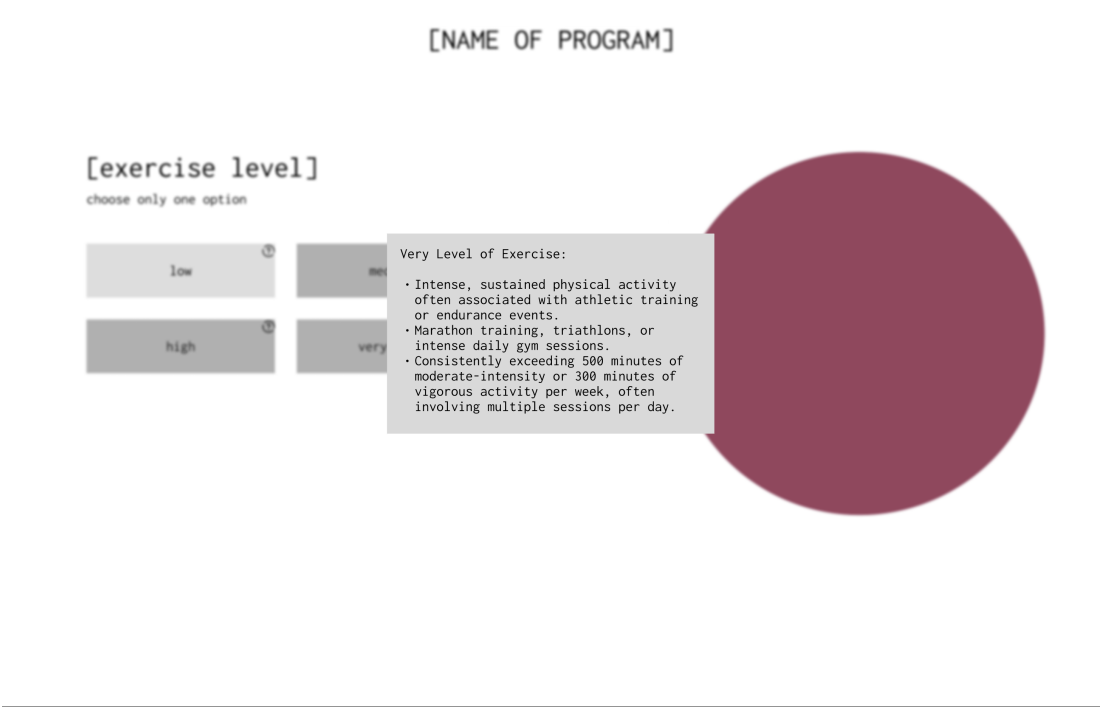


Figure B.5: An example of what occurred after the question mark in the corners of the choices was hovered over.

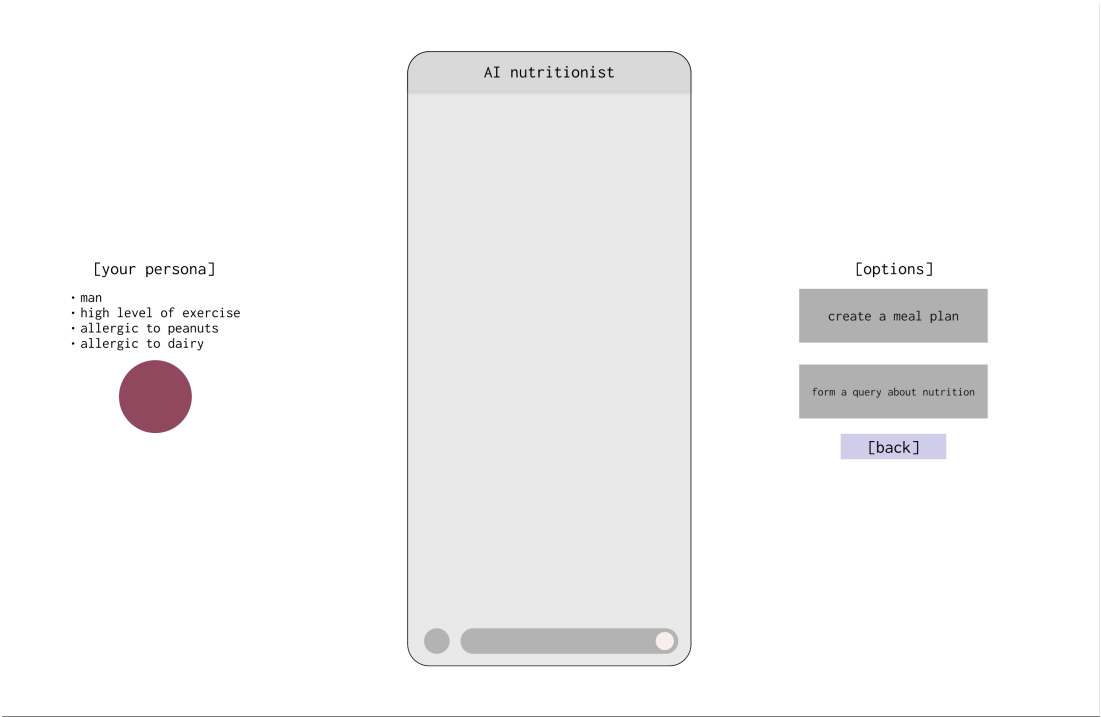


Figure B.6: The game interface of the Figma prototype. On the left-hand side: persona breakdown. On the right-hand side: query option examples.

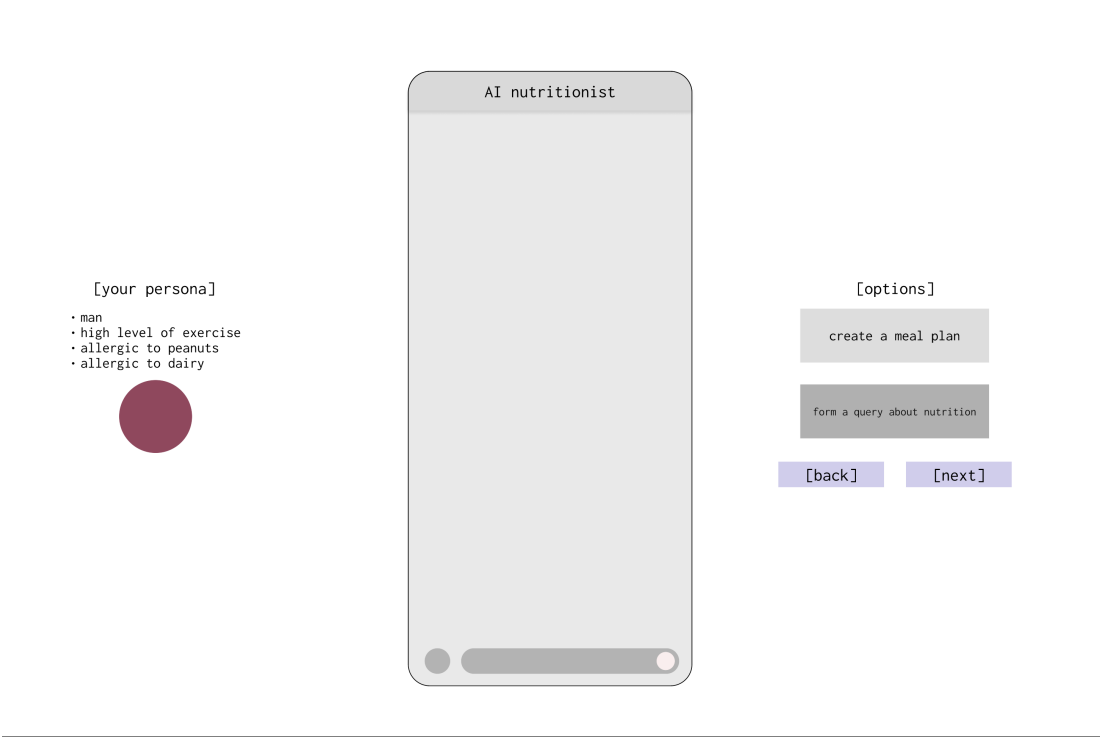


Figure B.7: The interface after an option has been selected.

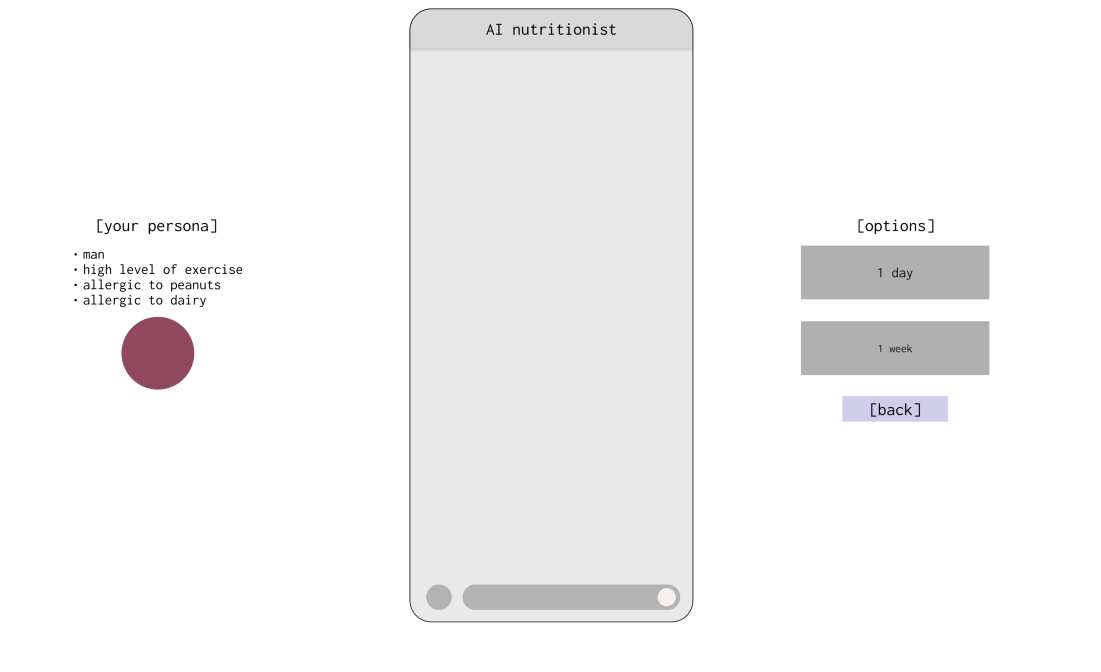


Figure B.8: The interface following the selection of “create a meal plan”.

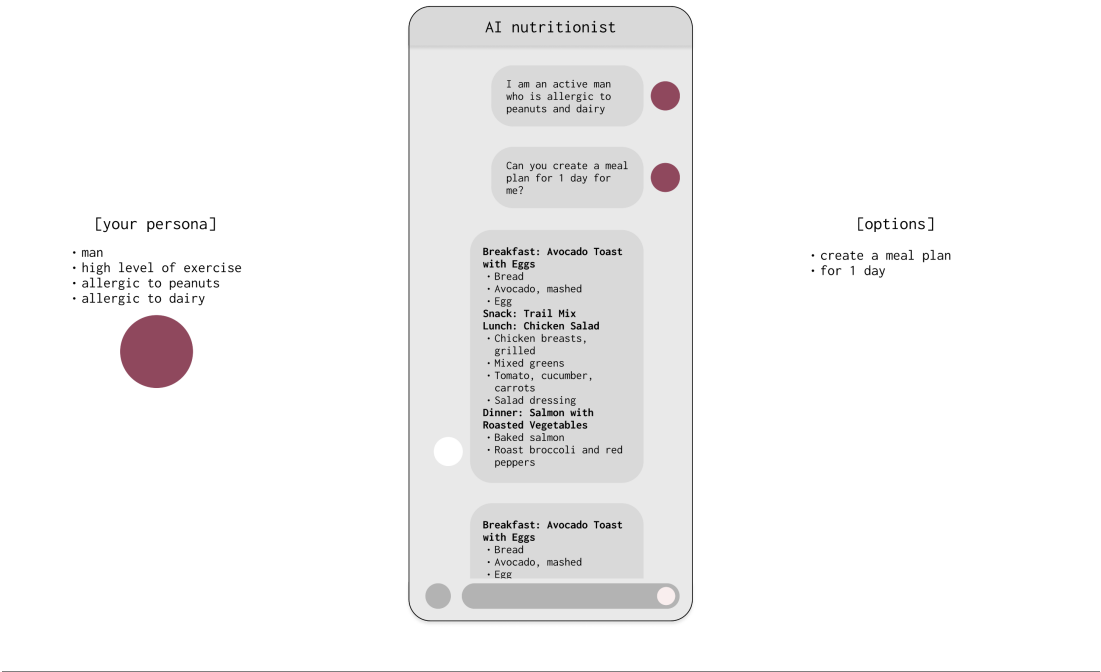


Figure B.9: The prototype interface after a query has been created.

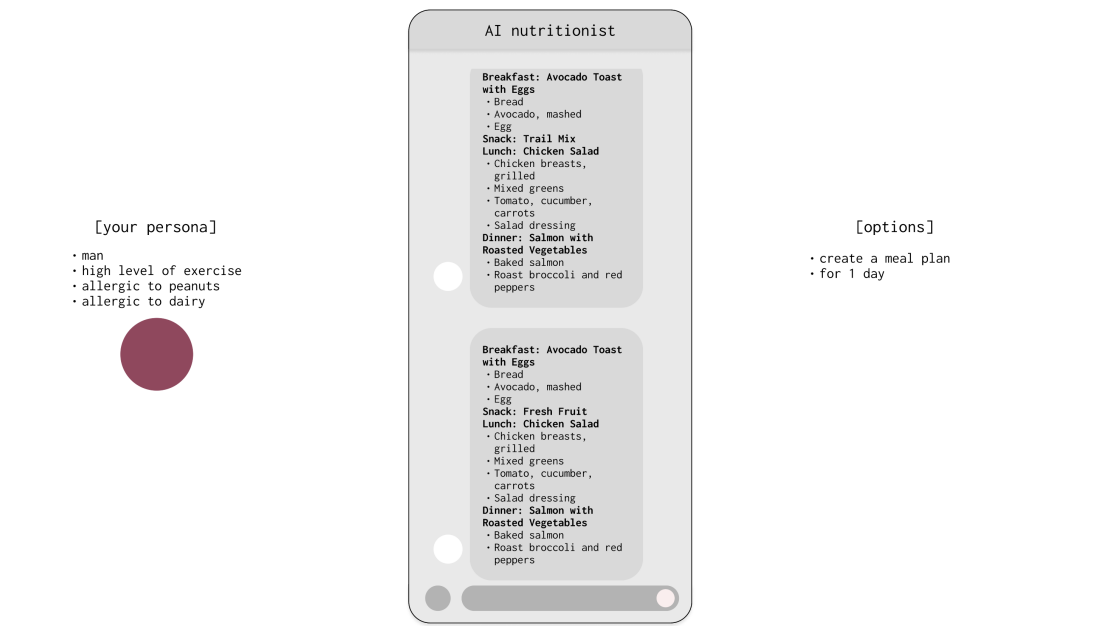


Figure B.10: The prototype interface after a query has been created. Scrolled down.

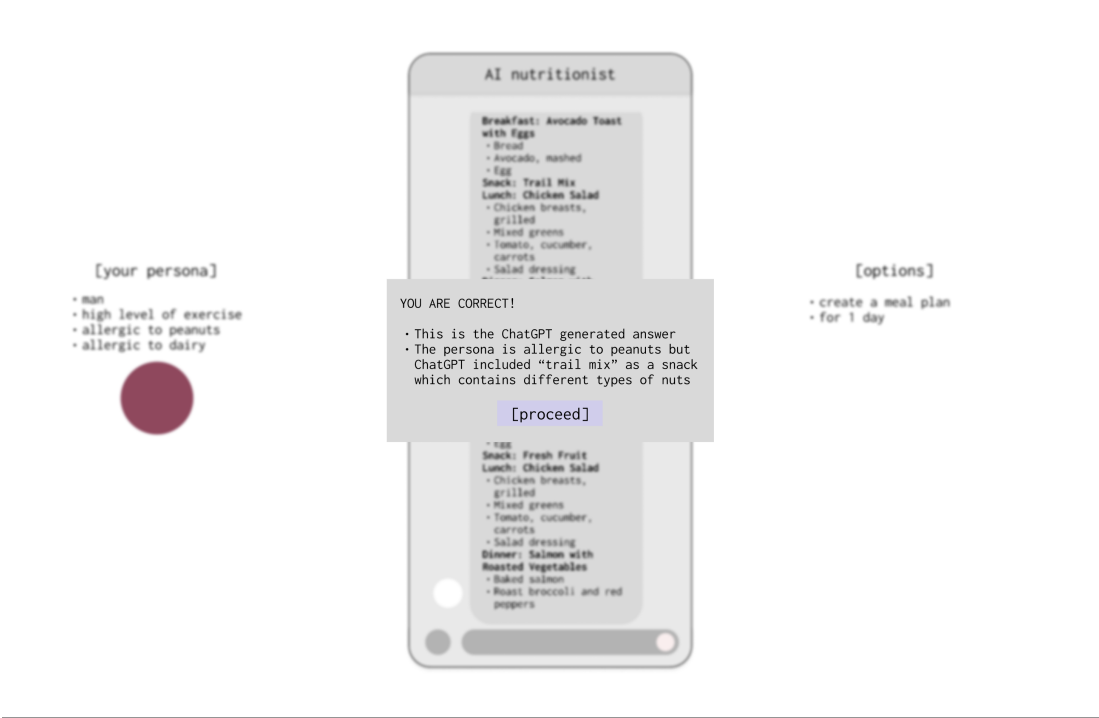


Figure B.11: The prototype interface after a response has been selected.



Figure B.12: The prototype interface after the pop-up has been closed.

Appendix C

Post-Focus Group 1 Functioning Prototype

The following images detail the interface of the functioning prototype developed after the initial focus group.

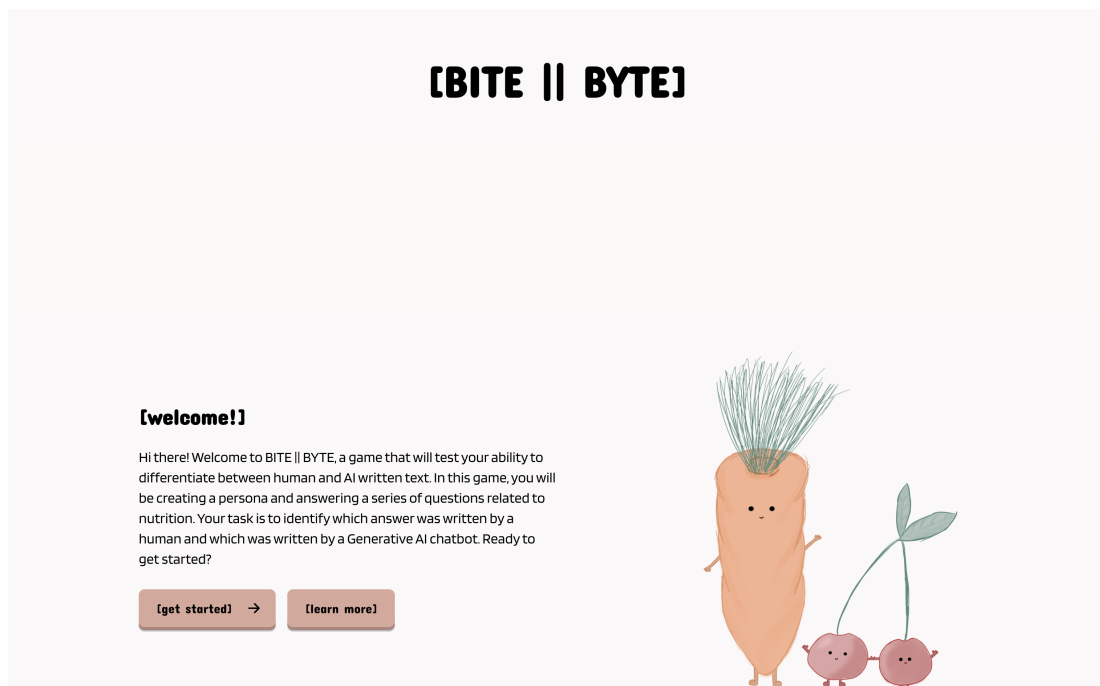


Figure C.1: The landing page of the functioning prototype, part 1.

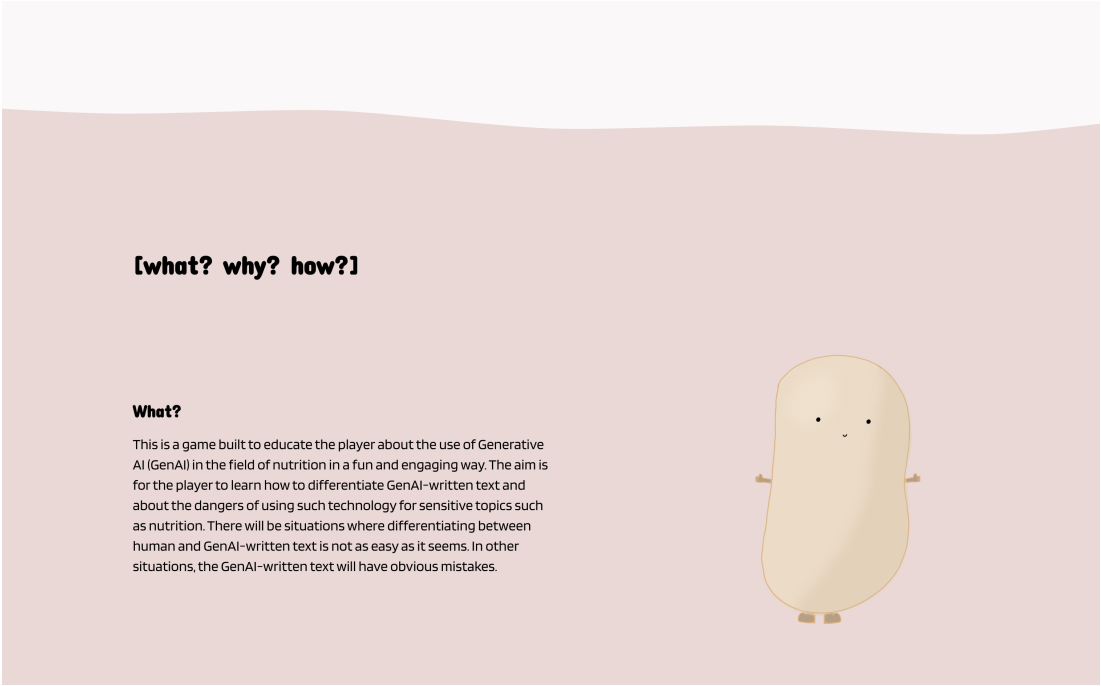


Figure C.2: The landing page, part 2.

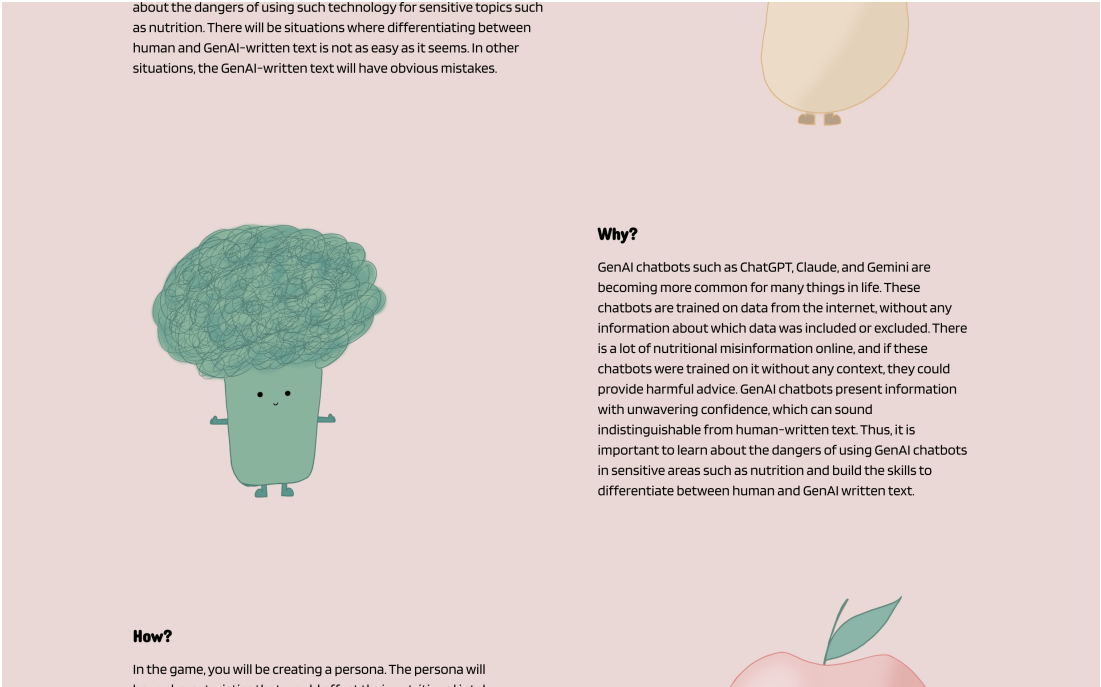


Figure C.3: The landing page, part 3.

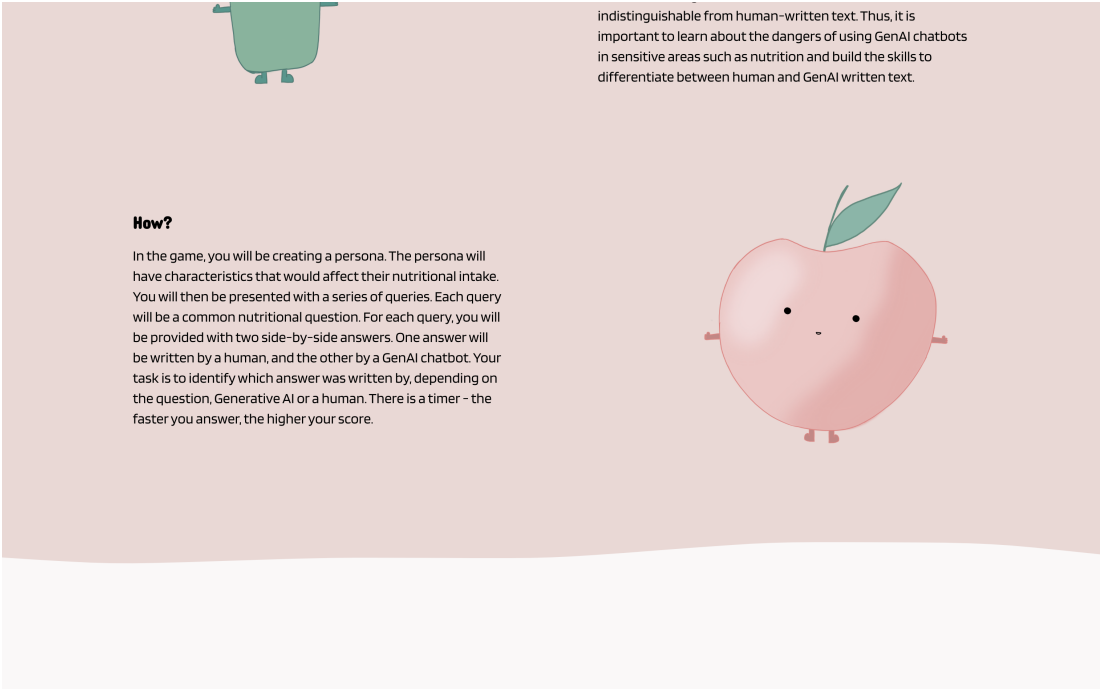


Figure C.4: The landing page, part 4.

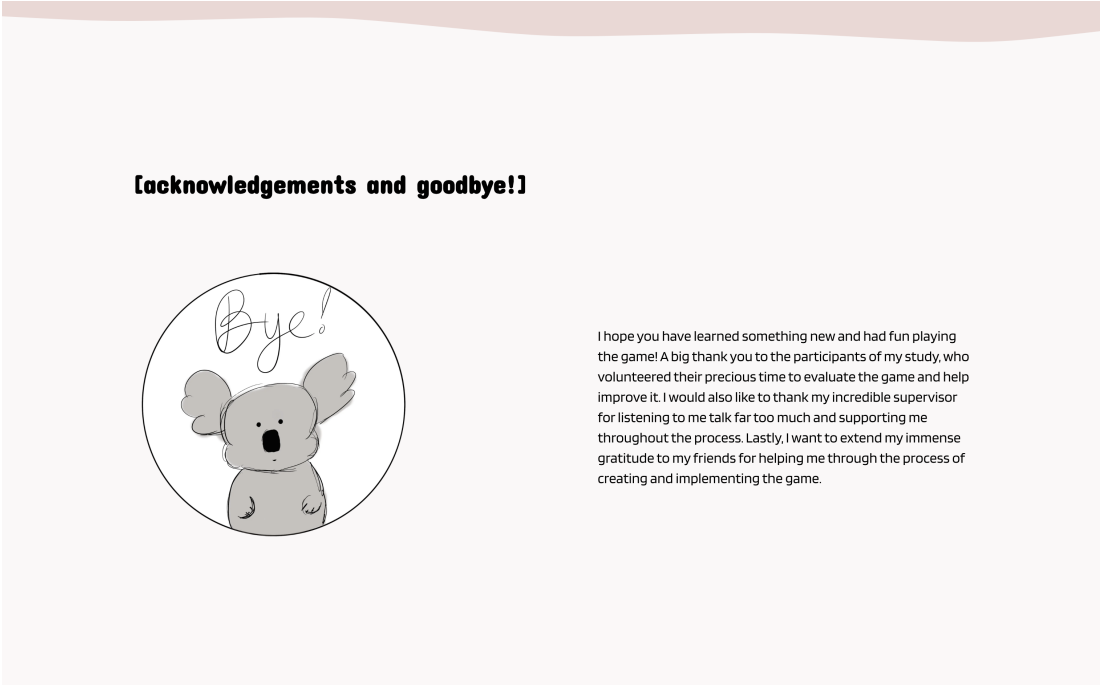


Figure C.5: The landing page, part 5.

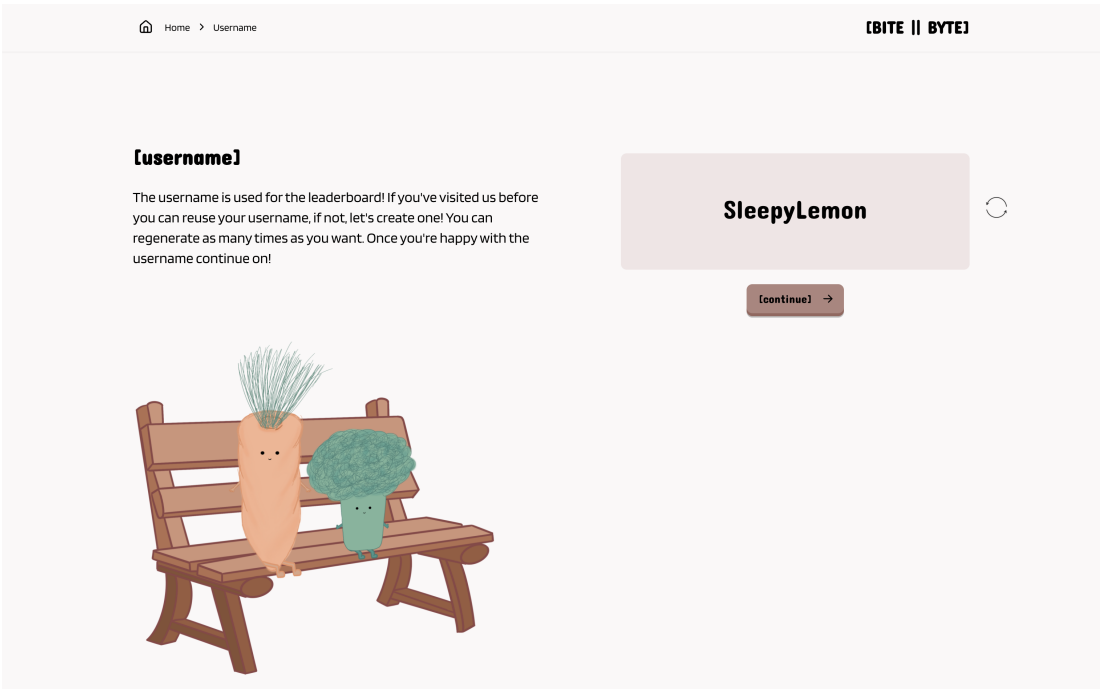


Figure C.6: The first page users see after selecting the “get started” button on the home page if they have not played before. This is the username generation page.

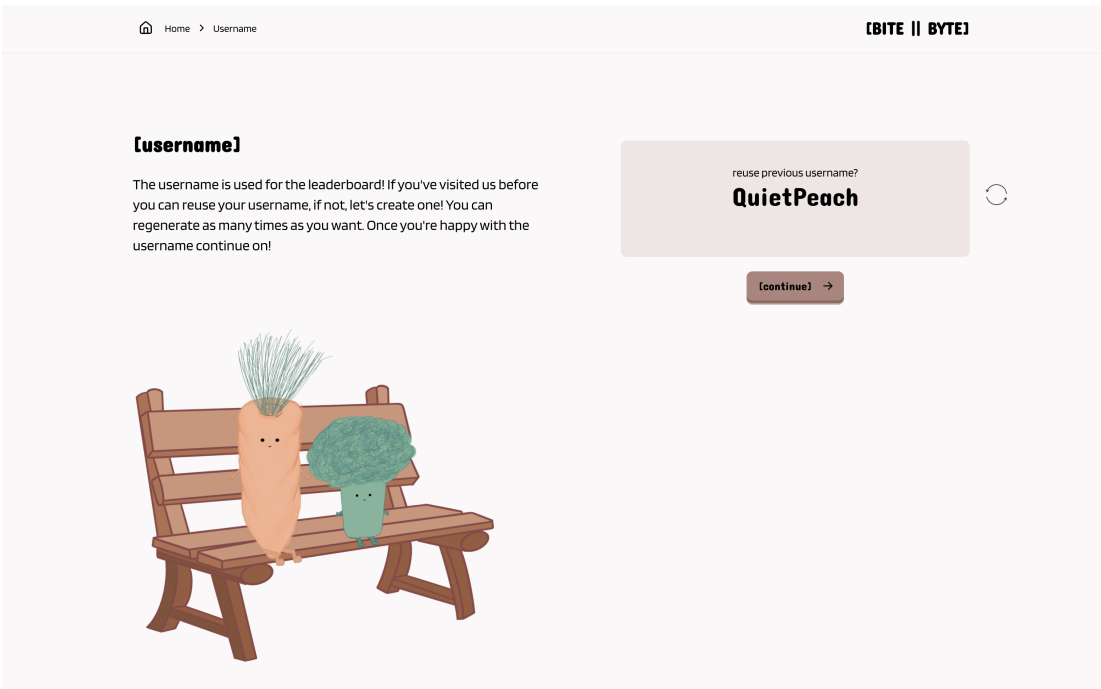


Figure C.7: The first page users see after selecting the “get started” button on the home page if they have played before. The option to reuse the users previous username is presented.

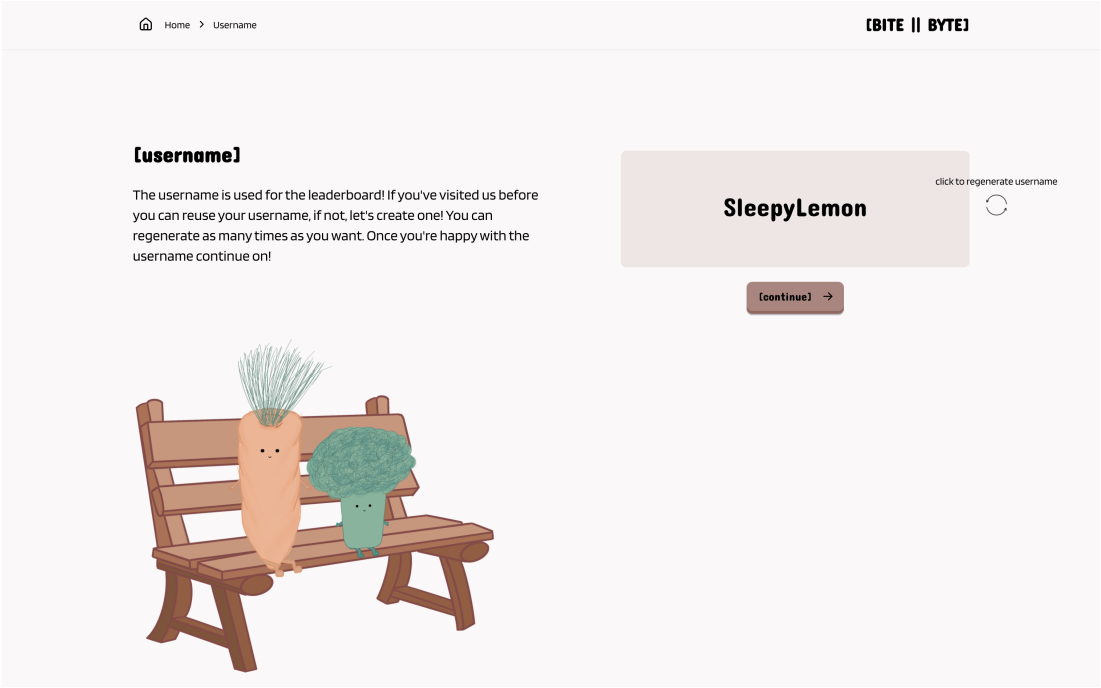


Figure C.8: The tooltip appears when a user hovers over the “regenerate” button.

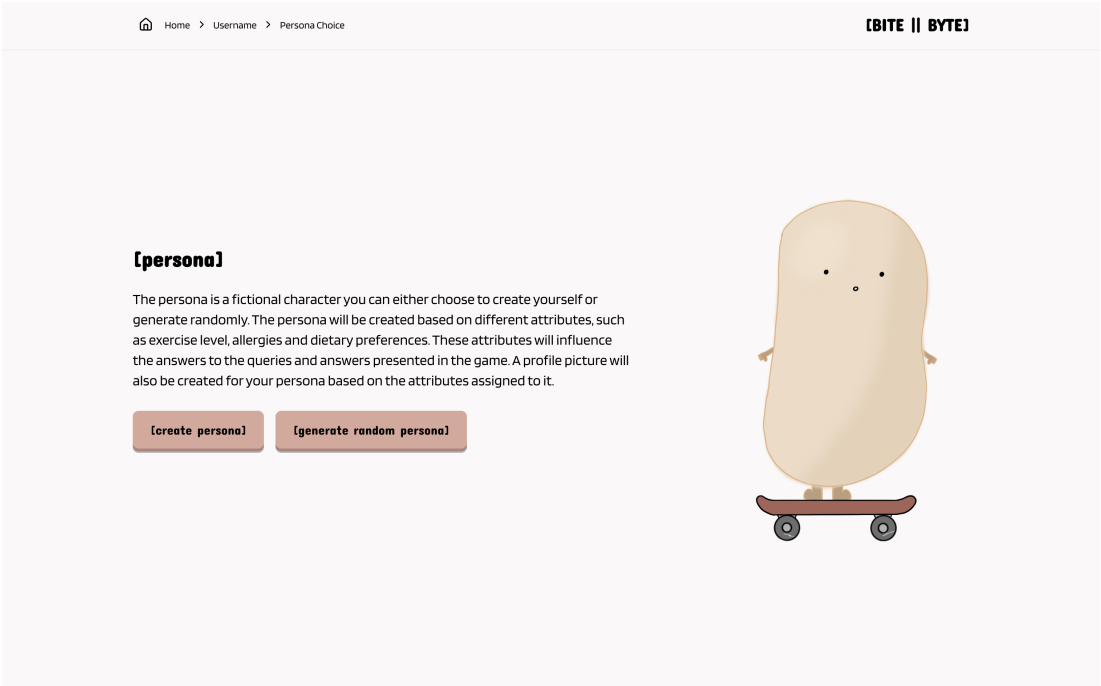


Figure C.9: This is the page following the username generation. The “generate persona” page does work, but as the game only functions with specific personas, the generated personas cannot be used.

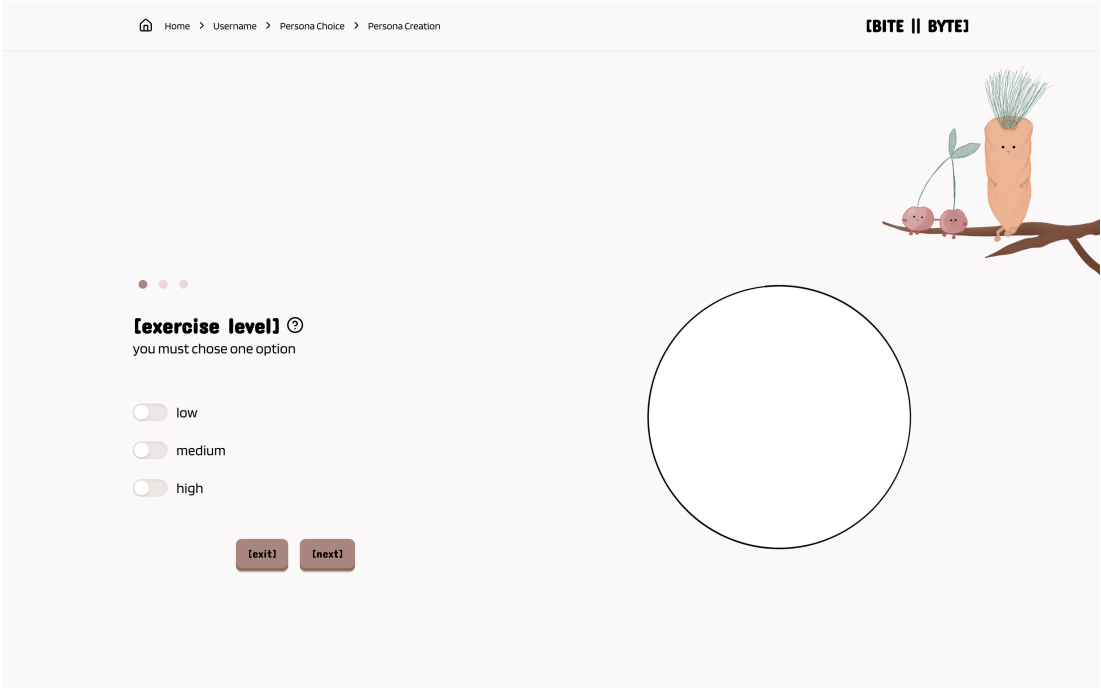


Figure C.10: This is the persona creation page.

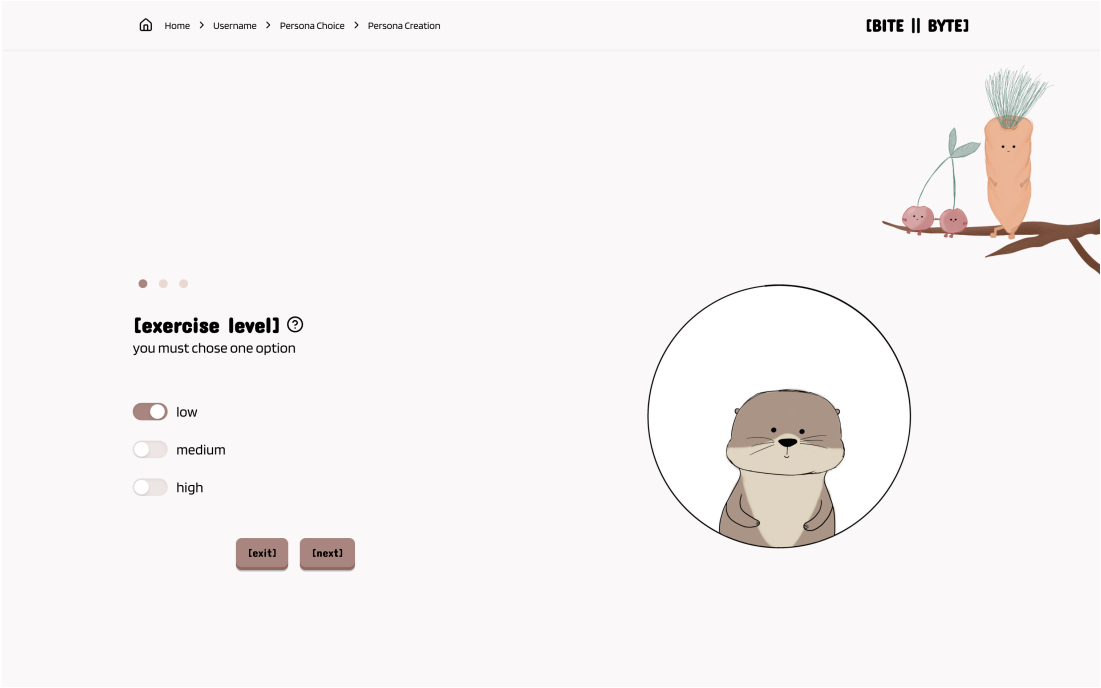


Figure C.11: An example of profile image generation after an exercise level (low) has been selected.

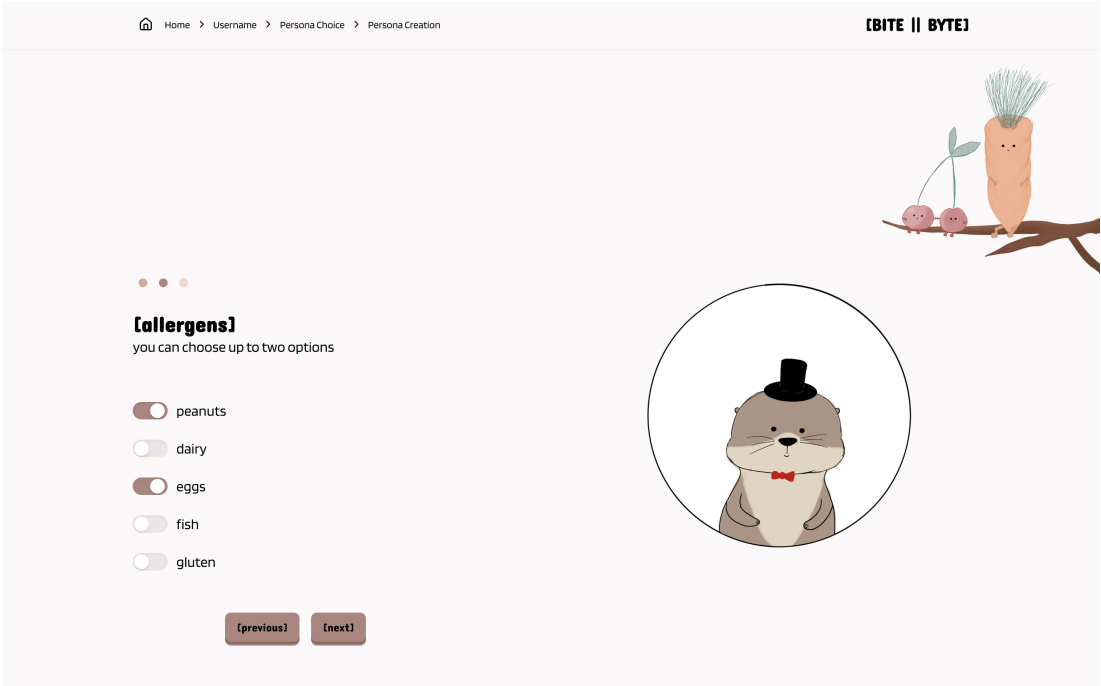


Figure C.12: An example of profile image generation after allergens (peanuts and eggs) have been selected.

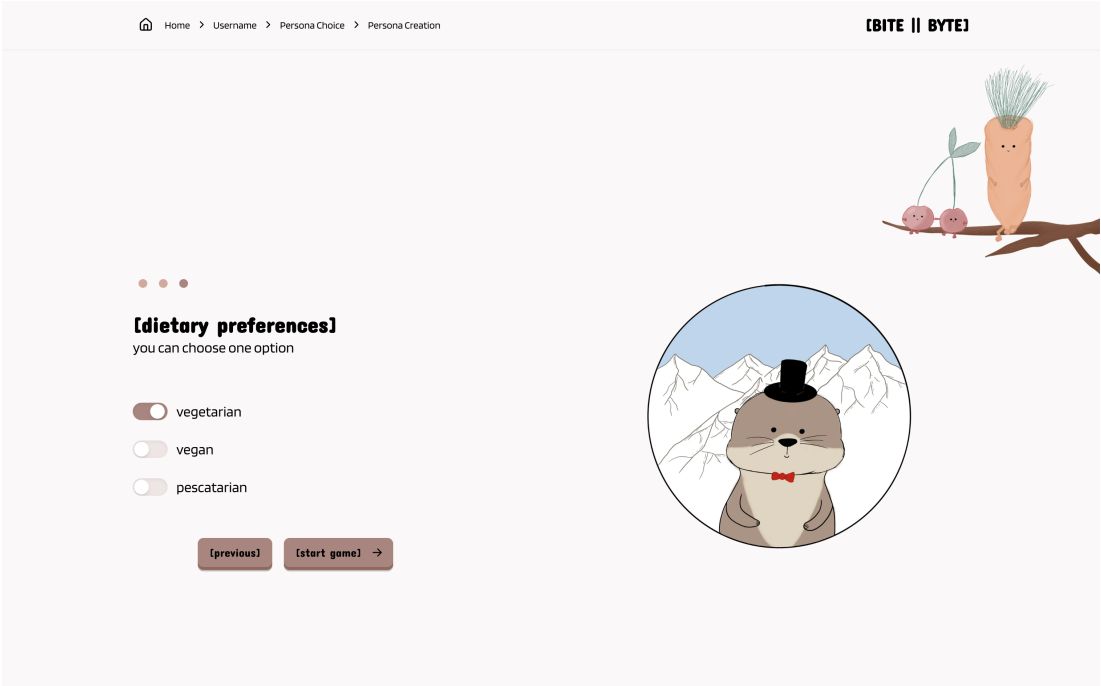


Figure C.13: An example of profile image generation after a dietary preference (vegetarian) has been selected.

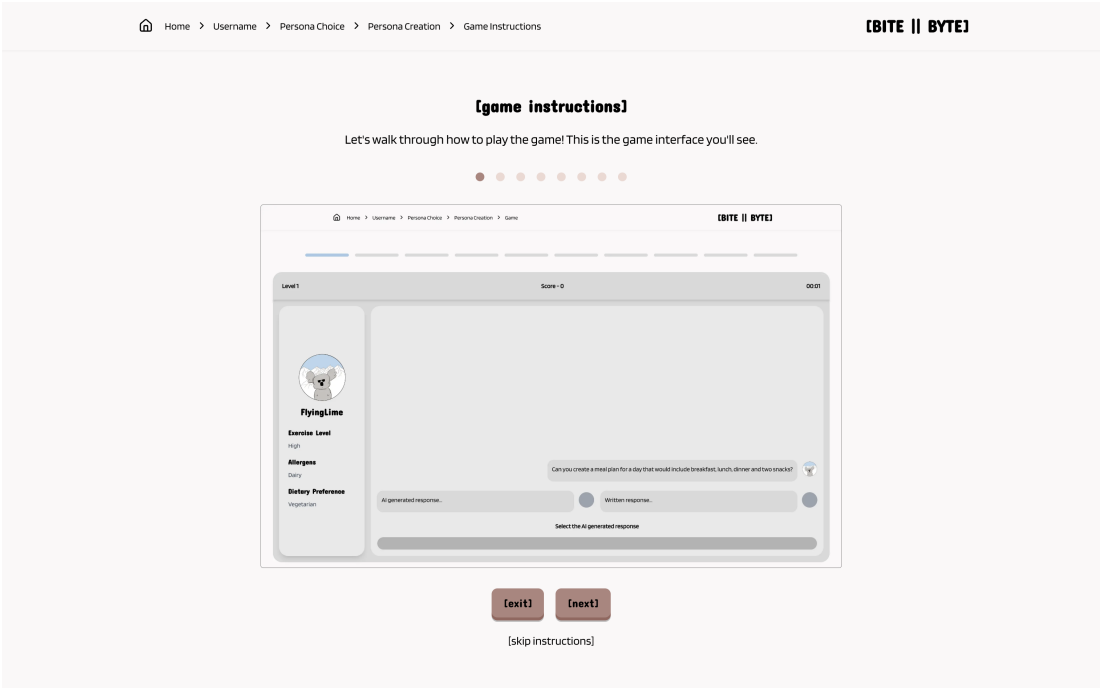


Figure C.14: This as an example from the instruction page.

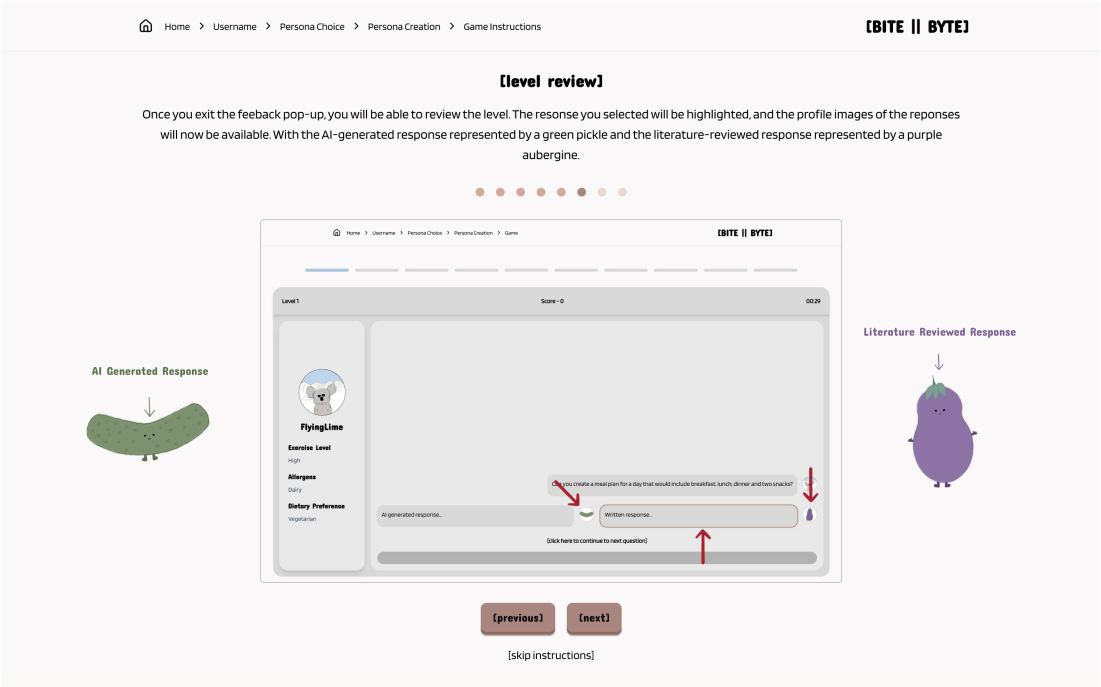


Figure C.15: Another example of an instruction page.

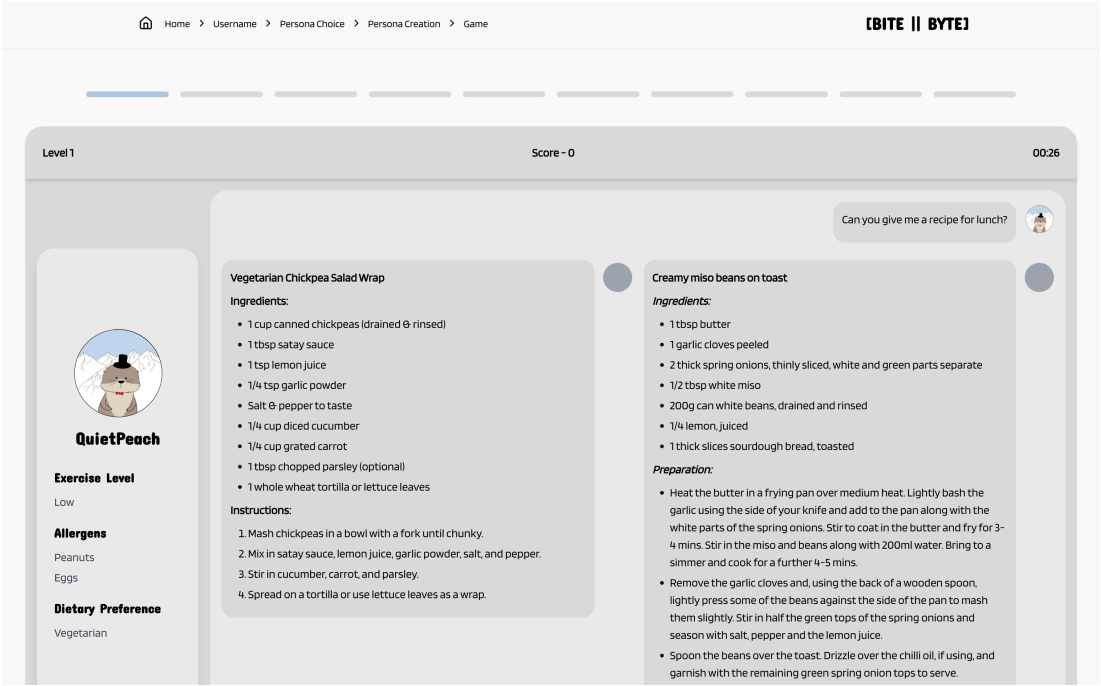


Figure C.16: A game page example for question 1 scrolled to the top of the page. For a vegetarian persona with low exercise level, a peanut and egg allergy.

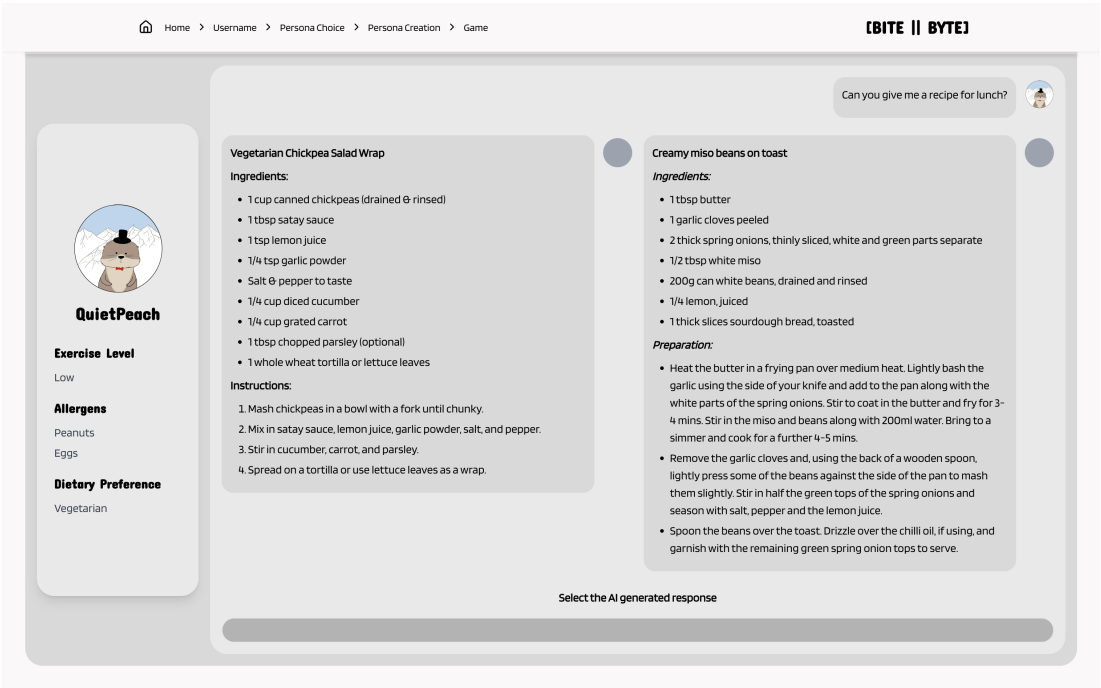


Figure C.17: A game page example for question 1 scrolled down to the bottom of the page.

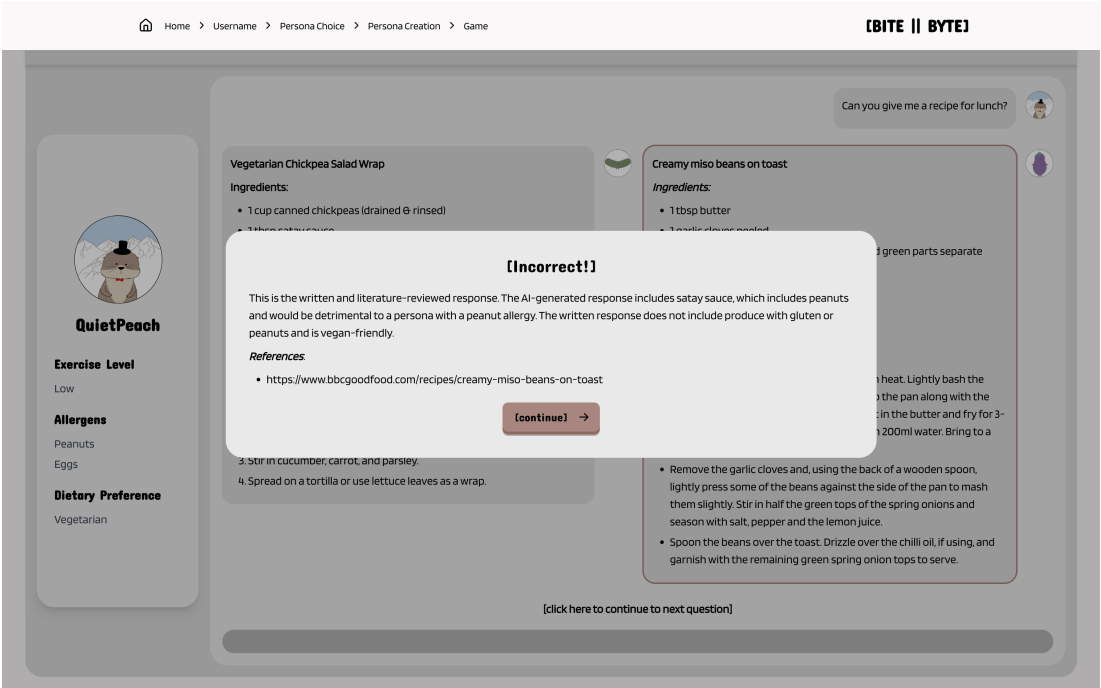


Figure C.18: An example of a feedback pop-up after question 1 has been answered incorrectly (the literature-reviewed response was selected).

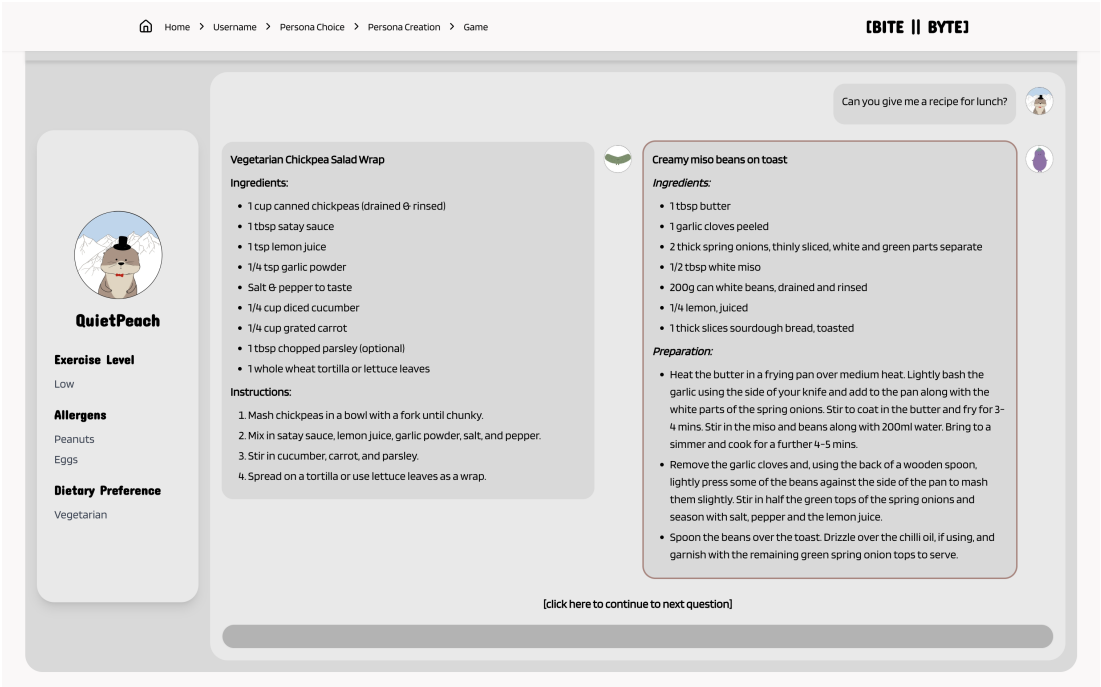


Figure C.19: A game page example after the pop-up has been closed.

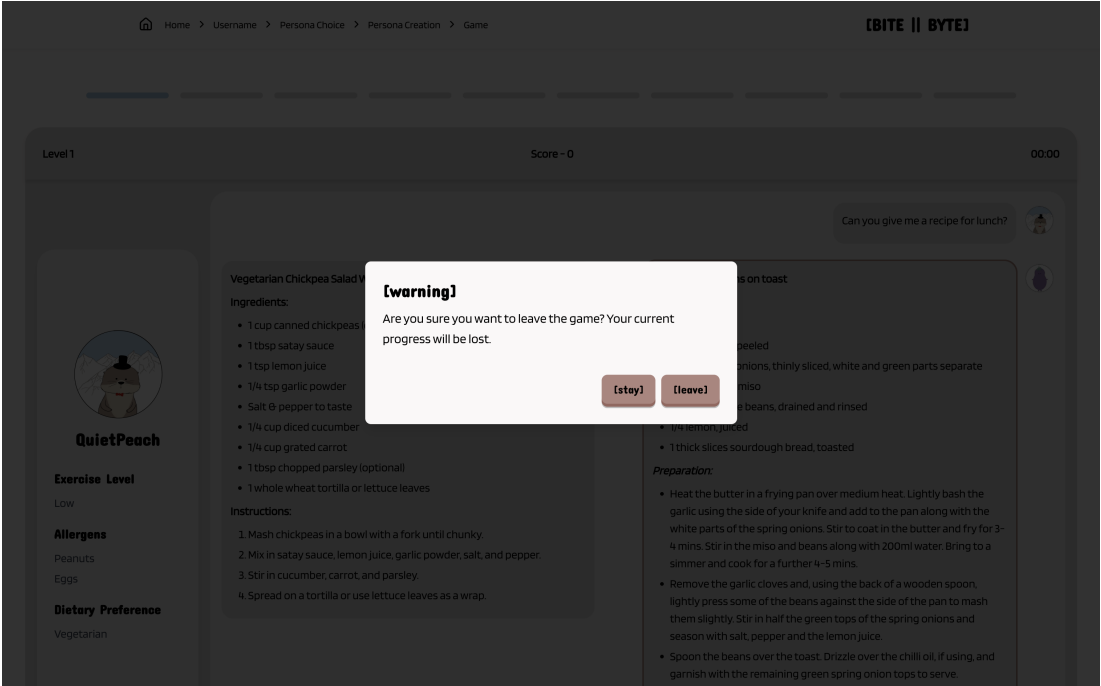


Figure C.20: The pop-up alert that appears when a user tries to navigate to other parts of the webpage when they are playing the game.

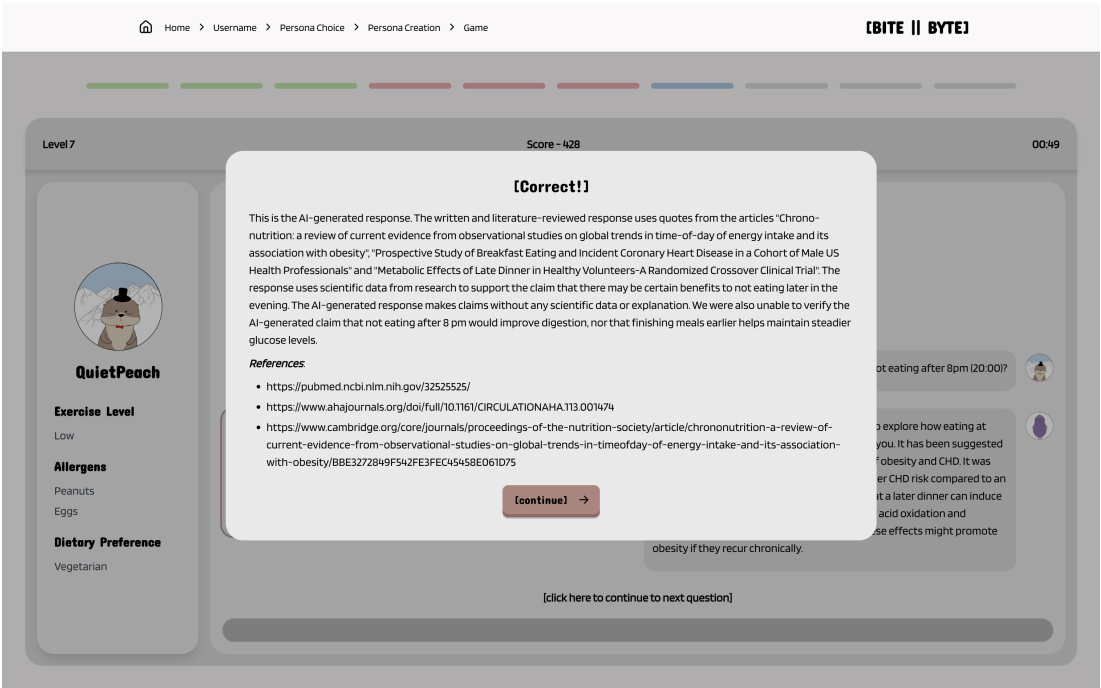


Figure C.21: A feedback pop example for a correct answer (the AI-generated response was selected). As this is question 7, the level progress bar can also be seen filling up.

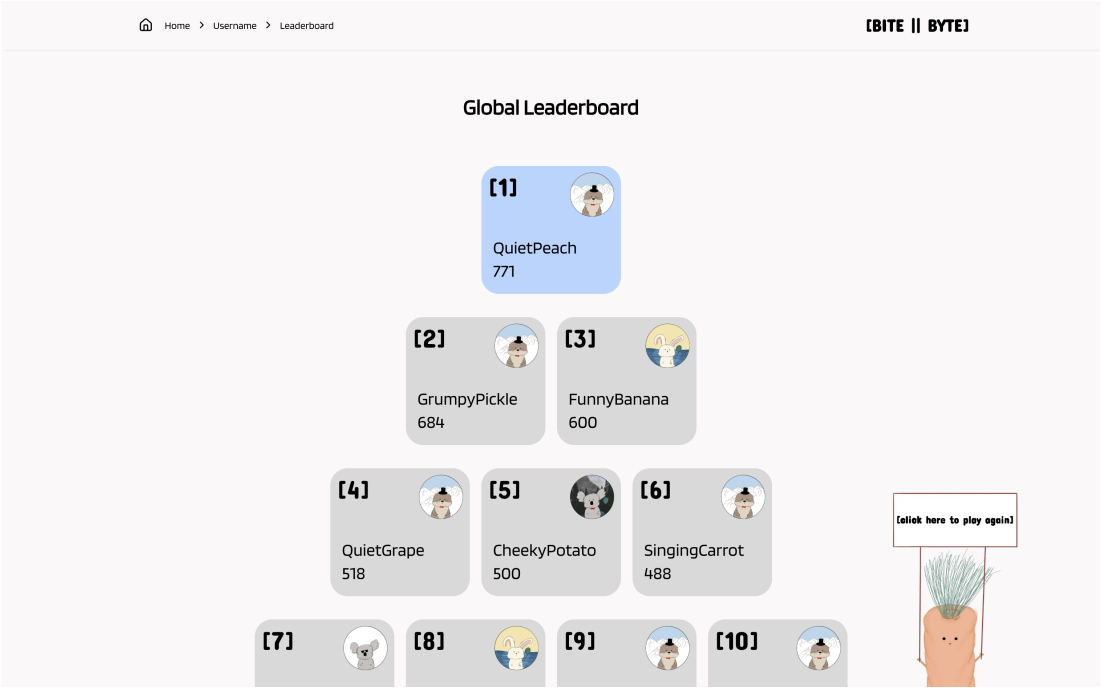


Figure C.22: An example of the leaderboard if the user makes it to the top 10 players.

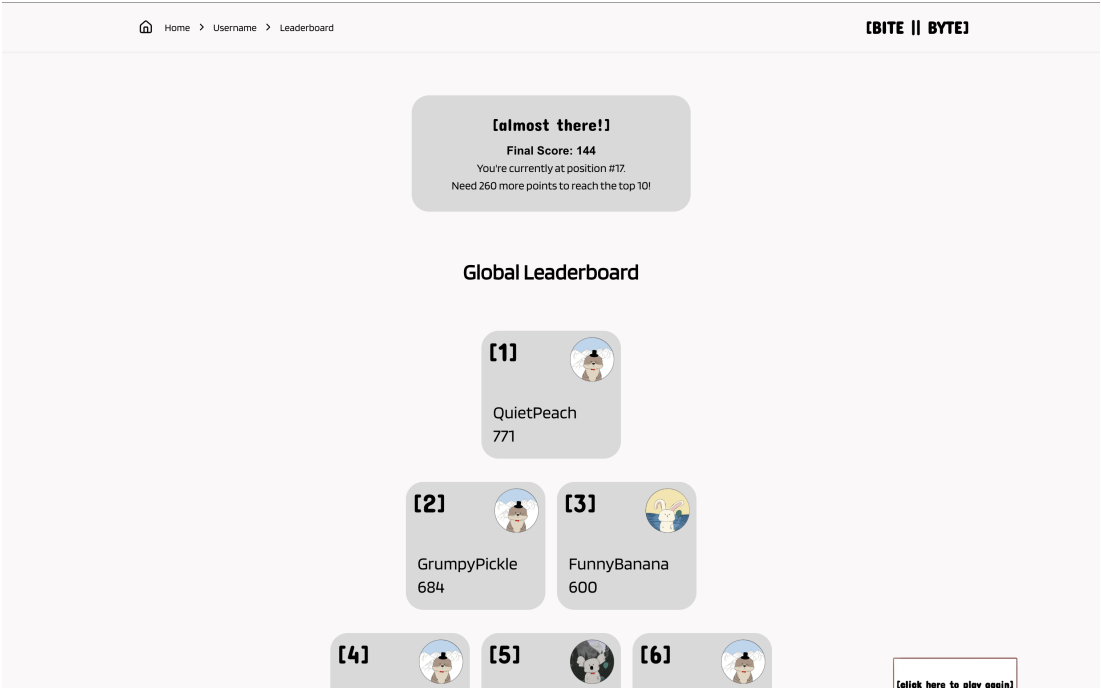


Figure C.23: An example of the leaderboard if the user does not make it to the top 10 players.

Appendix D

Final Functioning Prototype

The following images detail the interface of the functioning prototype changes made after the second focus group. These are the main visible changes to the interface. Some changes cannot be seen as they were changes in game mechanics, not the interface.

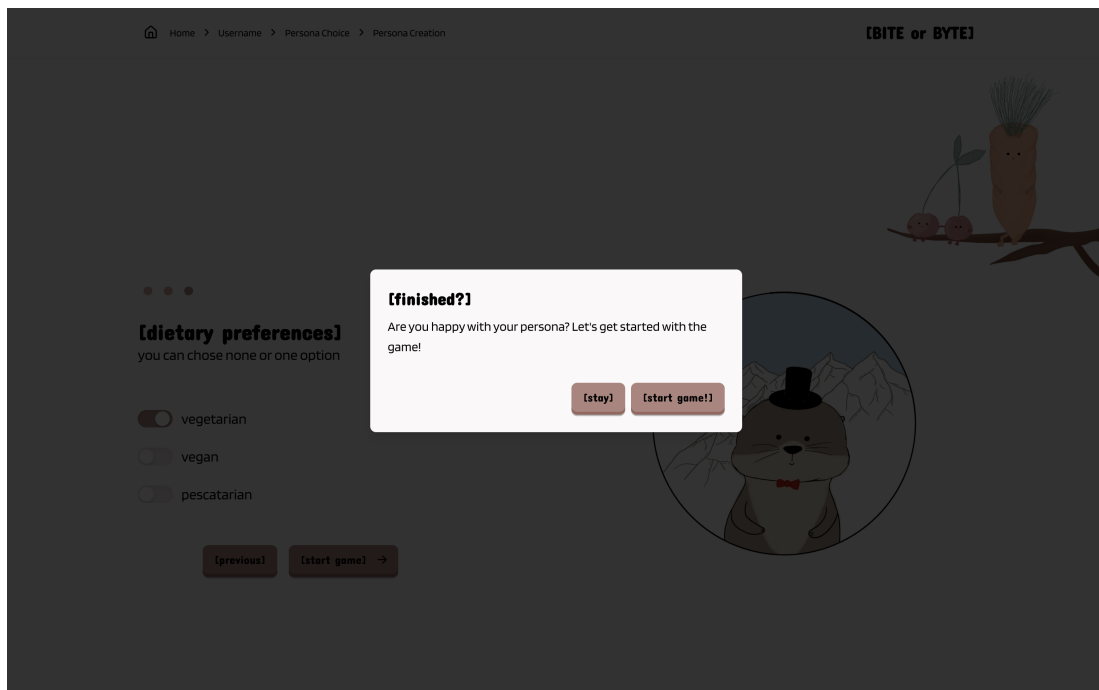


Figure D.1: This pop-up alert now appears when a user clicks the “start game” button after creating a persona. Previously, a user would be immediately directed to the instructions page.

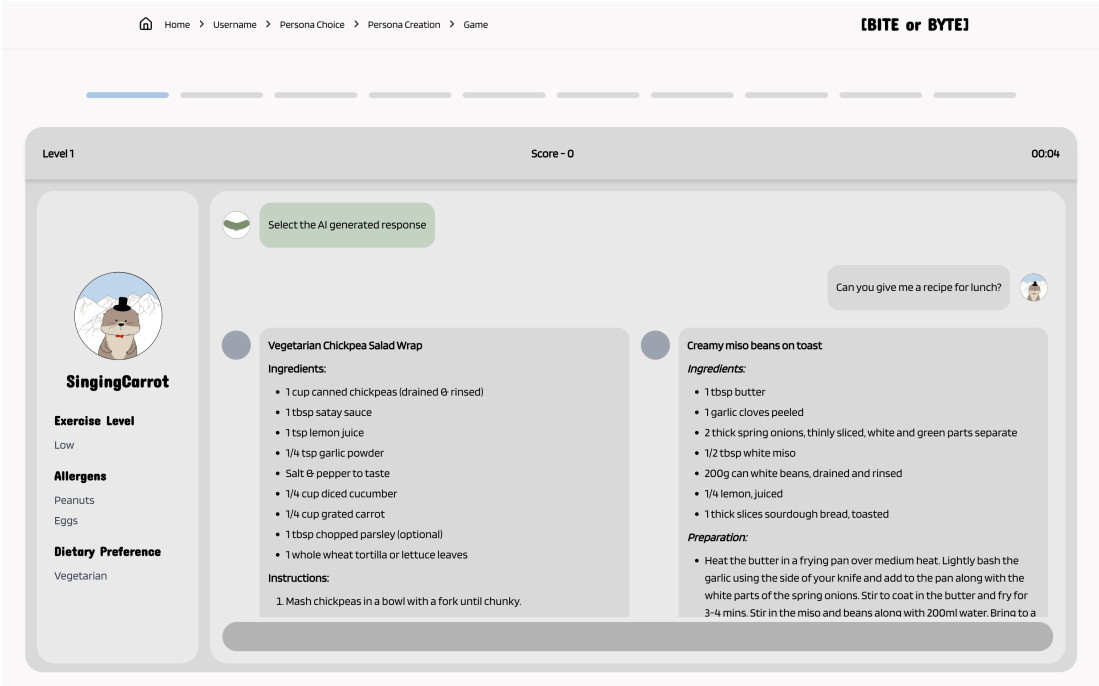


Figure D.2: This is the new game interface. Most elements are the same, except the messaging app-like interface is restricted to the view-port size with internal scrolling. The instruction to select the AI-generated response has also been moved to the top of the page.

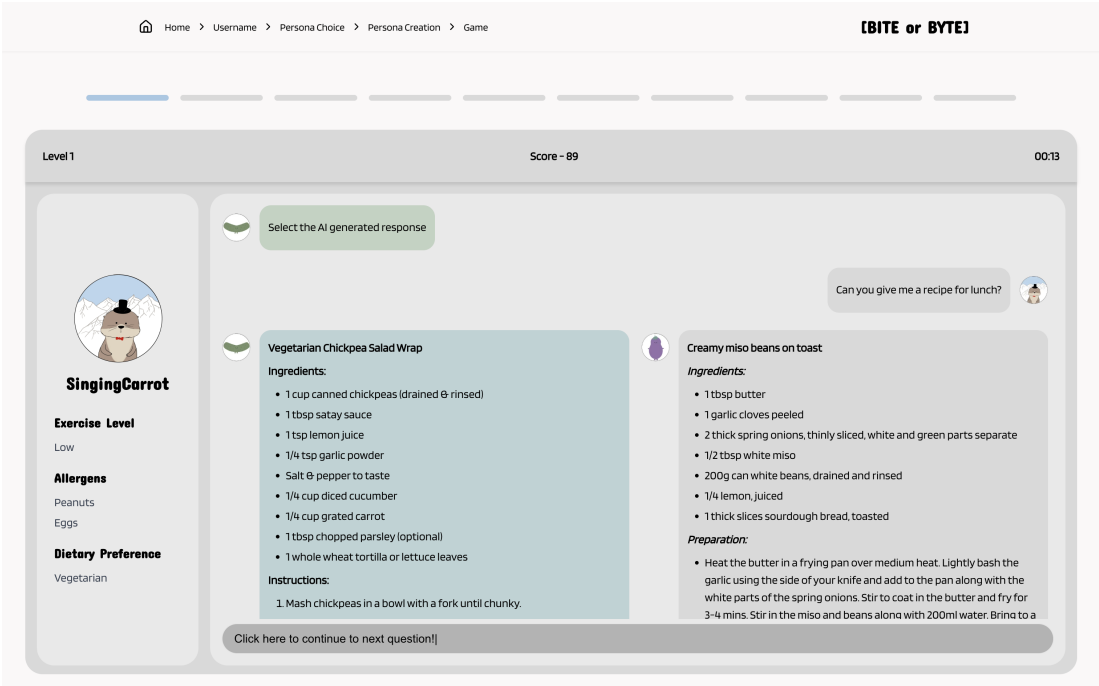


Figure D.3: This is the game interface after the feedback pop-up has been closed. The selected response is now highlighted more clearly and the button to proceed to the next level has been relocated to the bottom of the interface.

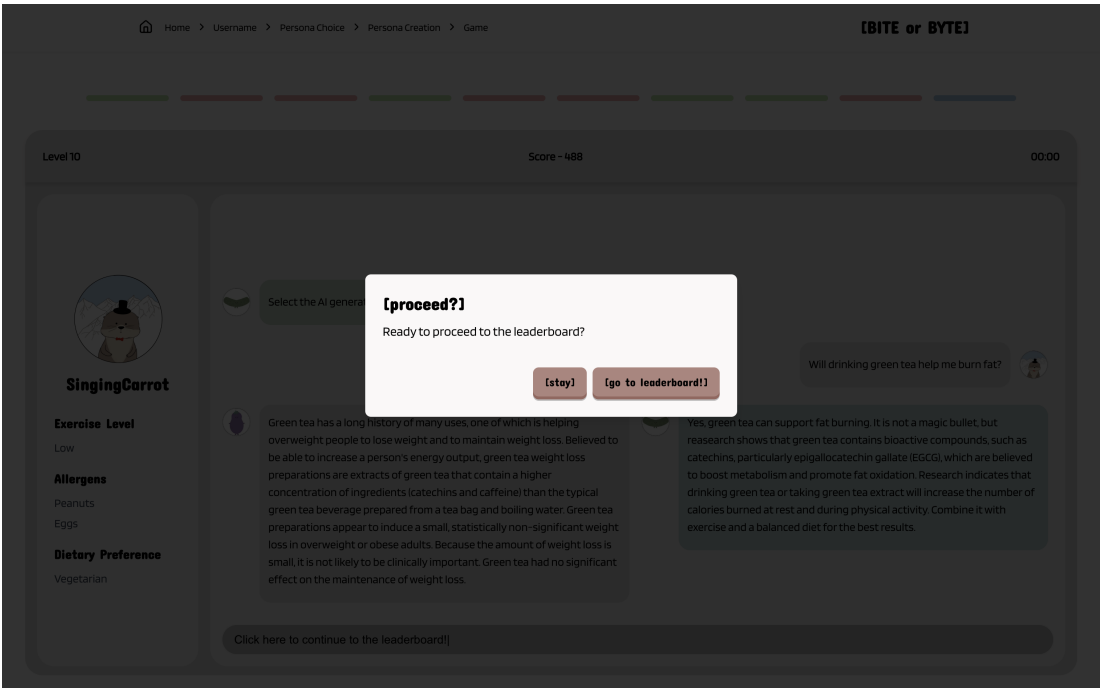


Figure D.4: This pop-up alert now appears when a user answers the final question and clicks the “continue to leaderboard” button. Previously, the user would be immediately directed to the leaderboard.

Appendix E

Analysis Across Focus Groups

E.1 Focus Group 1 Group Discussion Thematic Analysis

After the initial focus group, I performed a thematic analysis of the group discussion. I discovered the following themes around participants' use of GenAI chatbots (Figure E.1), participants' interactions with nutritional information online (Figure E.2), and participants' use of GenAI chatbots in nutritional contexts (Figure E.3).



Figure E.1: Themes surrounding participants' use of GenAI chatbots

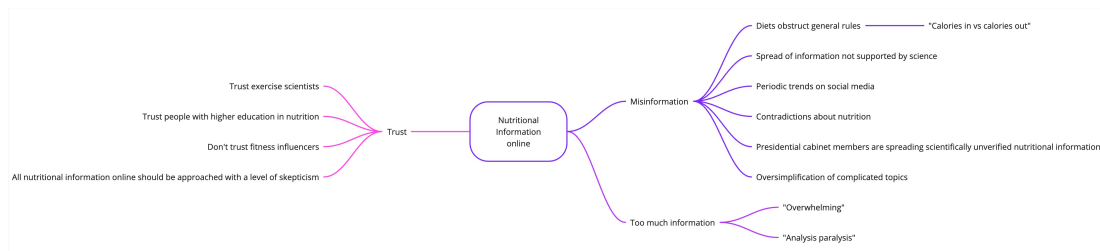


Figure E.2: Themes surrounding participants' interactions with nutritional information online.

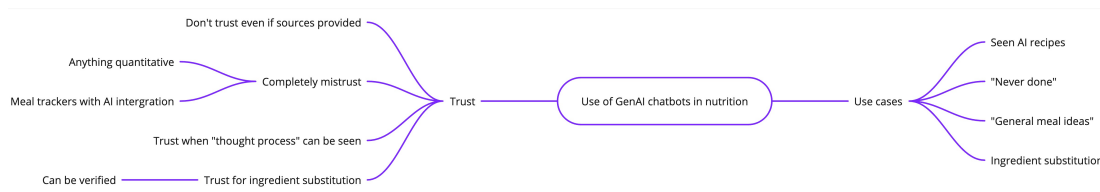


Figure E.3: Themes surrounding participants' use of GenAI chatbots in nutritional contexts.

E.2 Focus Group 2 Group Discussion and Individual Think-Aloud Thematic Analysis

After the second focus group, I performed a thematic analysis of the group discussion. I discovered the following themes: participants' opinions on the game interface (Figure E.4), participants' opinions on the game mechanics (Figure E.5), and the influence of the game on participants' use of GenAI chatbots (Figure E.6).

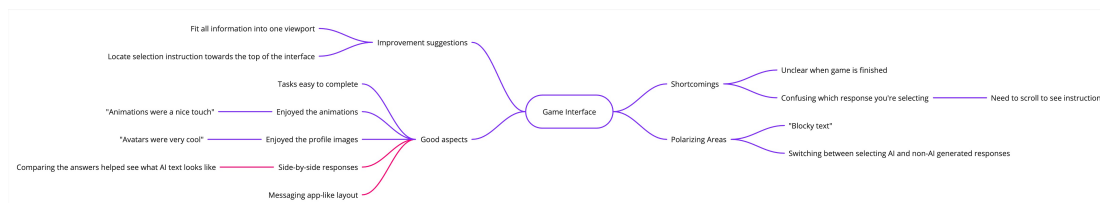


Figure E.4: Themes surrounding participants' opinions on the game interface.

E.3 Focus Group 2 Pain Point Identification

After the second focus group, I performed pain point identification using the recordings of the individual think-aloud sessions. Figure E.7 showcases the summary of the different pain points identified.

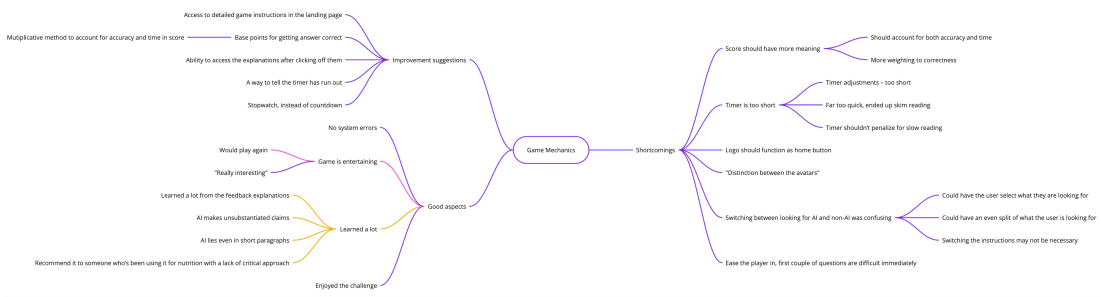


Figure E.5: Themes surrounding participants' opinions on the game mechanics.



Figure E.6: Themes surrounding the influence of the game on participants' use of GenAI chatbots.

Category	Description	User Impact	No. of participants effected	Overall Frequency
Content	Persona creation characteristic constraints unclear	Low	1	1 instance
UI/UX	Task completion uncertainty	Low	3	6 instances
UI/UX	User did not realize the timer had run out	High	3	21 instances
Technical	Cannot reopen feedback pop-up	Medium	1	1 instance
UI/UX	Navigation to username creation instead of persona when replaying	Low	1	1 instance
UI/UX	Logo does not navigate to home page	Medium	1	1 instance
UI/UX	User did not see the change in response instruction	High	3	3 instances
Content	Use of usernames in leaderboard unclear	Low	1	1 instance
UI/UX	Username generation confusion	Low	1	1 instance
Technical	Timer too short	High	1	10 instances
Technical	No functionality for clicking the reference	Medium	1	1 instance
Technical	Score meaning unclear	Medium	2	2 instances
UI/UX	Too much text	Low	1	1 instance

Figure E.7: Summary of identified pain points.