Retrofitted: Enhancing Immersion in SoloRPGs through Large Language Models

Matthew Jordan

July 2024

Creative Technology

Faculty of Electrical Engineering, Mathematics, and Computer Science (EEMCS)

University of Twente

Supervisor: Marcus Gerhold

Critical Observer: Thérèse Bergsma

Abstract

This thesis explores enhancing player immersion in Solo Role-Playing Games (SoloRPGs) through integrating Large Language Models (LLMs). Traditional SoloRPGs, while offering unparalleled freedom and creativity, often suffer from complex rules and setups that can hinder player immersion. Attempts to transform SoloRPGs into digital formats have had some success but at the cost of reducing player choice due to pre-coded limitations. This project investigates whether LLMs, known for their human-like conversational capabilities and ability to adapt and respond to a wide variety of scenarios, can bridge the gap and provide a solution that enhances immersion while maintaining player freedom.

The research is guided by the primary question: "To what extent can player immersion in a SoloRPG be enhanced by a natural language interface using large language models?" Supporting this inquiry are sub-questions defining player immersion, the techniques to measure it, and applying LLMs in creating natural language interfaces. The methodology involved creating a modified version of the SoloRPG "Quill," incorporating OpenAI's GPT-3.5 Turbo model. This version was tested against the traditional game through user evaluations that measured flow, presence, and cognitive absorption using the Game Engagement Questionnaire.

Findings from 36 participants indicated no significant difference in flow and presence between the traditional and modified versions. However, a significant increase in cognitive absorption was observed in the LLM-enhanced version, suggesting that natural language interfaces can indeed enhance certain aspects of immersion in SoloRPGs. The study acknowledges limitations such as the scope restricted to a single game and potential biases introduced by the digital interface. Future research directions include expanding the range of SoloRPGs tested, exploring personalized AI responses, and leveraging advancements in AI to enrich interactive storytelling experiences further.

Contents

Chapter 1. Introduction

This introduction will provide an overview of the project's context, highlight the problem statement, and define the primary- and sub-research questions.

1.1 Context

Solo role-playing games (SoloRPGs) are a creative approach to a single-player experience where the player is only limited by their imagination and ability to immerse themselves into the world. SoloRPGs occupy a unique niche within the broader RPG genre, standing out with their single-player format that fosters a profoundly personal interaction with the game's story and mechanics. These games, often played through tabletop setups or narrative-driven books, offer a blend of rich storytelling and strategic decision-making. Their allure frequently lies in their ability to ignite a player's imagination, enabling them to visualise and interact with a complex fantasy world. Players embark on quests, unravel puzzles, and navigate through intricate plots, all while managing game elements like character progression, inventory, and event outcomes typically governed by dice rolls and a set of predetermined rules. While SoloRPGs provide the most freedom of any game medium, various barriers can diminish players' experience by preventing them from fully immersing themselves. As SoloRPGs can require players to perform actions such as rolling dice, referencing guides and tables, taking notes, and playing with complicated rulesets, players may find it challenging to maintain concentration and continuous immersion in the game world whilst juggling everything else. With many pieces, rules, and considerations, some of which can be seen in the setup for the game "Maquis" in Figure 1, may diminish player experience.

Figure 1: Setup of a Round of Maquis

This problem has already been tackled by implementing SoloRPGs into a digital game format. By switching the medium of play to digital, elements such as random number generation, looking up guides, and basic rules are automated. Thus, players can concentrate more on aspects of the game that involve storytelling and decision-making, allowing them to immerse themselves in the world better. However, by altering SoloRPGs to be played digitally, certain aspects of gameplay are lost. In contrast to playing a traditional SoloRPG, choices in a computer game must be predefined as these games run on prewritten code. For example, if a player wished to walk around a town and gather information from the local people and this was not coded into the game, that option would be

completely unavailable to them. Thus, digital implementations of SoloRPGs limit the players' freedom, constraining one of the genre's defining features.

With the recent innovation of Large Language Models (LLMs) such as OpenAI's GPT series, a solution that allows player freedom whilst enabling them to immerse themselves fully may be possible. LLMs are a type of artificial intelligence that uses massive text databases to replicate human-like speech and can be utilised to create natural language interfaces capable of accepting a wide range of inputs. These models leverage deep learning techniques to process and understand the intricacies of human language, making them adept at generating coherent and contextually relevant responses to a wide variety of situations. This capability may lend itself to interactive applications like SoloRPGs, where the narrative can evolve dynamically based on the player's input. A possible advantage of using LLMs in SoloRPGs is their ability to maintain the fluidity and open-mindedness of traditional role-playing games. Unlike predefined digital games that restrict player choices to programmed options, LLMs can generate responses and scenarios on the fly, allowing players to explore a limitless array of possibilities. This enhances the sense of freedom and ensures that the game remains engaging and unpredictable, closely mimicking the experience of playing with a human game master. As a result, it may be possible for players to immerse themselves more deeply into the game world.

1.2 Problem Statement

For this project, the challenge will be to create a SoloRPG experience that uses a natural language interface to allow players to maintain their freedom within the game while enhancing immersion. Large language models (LLMs) will be used to create this natural language interface.

1.3 Research Questions

To address this project challenge, the central research question of this project was developed:

RQ: "To what extent can player immersion of a SoloRPG be enhanced by a natural language interface using large language models?"

As the main research question covers a broad area, several sub-research questions were developed to help answer it and guide research:

Sub RQ1: "What is player immersion?"

Sub RQ2: "What are the techniques used to measure player immersion?"

Sub RQ3: "How are large language models used to create natural language interfaces?"

1.4 Report Structure

This report is structured in a series of chapters that aim to explain the process of developing and evaluating the AI-enhanced SoloRPG. This chapter serves as an introduction to the project context and an explanation of the challenge and research questions. The next chapter contains background research that includes a literature review, a state-of-the-art, and a conclusion that describes the current landscape of this project. Chapter three describes the methods and techniques that will be used during the project. Chapter four covers the ideation phase, in which various analyses are performed, preliminary requirements are formed, and initial concepts are conceptualised and selected for further development. Next, in chapter five, specifications are finalised using personas, the creation of Lo-Fi prototypes and testing. In chapter six, the realisation of the final product is detailed, explaining which tools and methods were used. The next chapter covers the evaluation of the product, including a description of the results and their analysis. In chapter eight, the project is concluded upon, and limitations are noted and discussed. Finally, chapter nine discusses further work that may be done towards the topic of this paper.

Chapter 2. Background Research

2.1 Literature Review

2.1.1 Defining Immersion

Before delving into the methods of measuring player immersion, it's crucial to establish a clear understanding of what exactly is being measured. Player immersion in games is a complex concept that has been extensively explored in academic literature. However, despite the wealth of research, there is no unanimous definition of immersion in games. Three key terms - Flow, Presence, and Cognitive Absorption - are often used interchangeably with immersion, each describing a unique aspect of the overall concept. [1] [2] [3]. These three terms all describe a form of immersion and are similar; however, an overarching concept of immersion may be concluded by gaining an in-depth understanding of each term and their differences.

Flow is the first of these concepts and was initially described by Csikszentmihalyi in 1990 as a positive state of mind in which “individuals are so involved in an activity that nothing else seems to matter” [4]. Beyond this initial definition, current literature identifies flow as a sense of enjoyment and deep involvement felt in gaming action [5] [6] that occurs when a player feels as though they are performing “well and effortlessly” [7], which is achieved by a balance of challenge and individual player skill [5]. Hence, flow is a sense of deep engagement that stems from a balance of player skill and challenge.

The next concept is presence, which refers to the sensation of being a ‘part’ of a virtual environment. The term was first introduced in a questionnaire developed by Witmer and Singer in 1998, in which they defined presence as “the subjective experience of being in one place or environment, even when one is physically situated in another” [8]. In his explication of presence, Lee [10] identifies three forms of presence. Self-presence refers to a state in which users experience their virtual selves as their actual selves, and social presence refers to a state in which users experience virtual social interactions as real [9] [10]. Jin further broke down the last form Lee identified [11] into physical and spatial presences according to differing definitions within literature. In this model of presence, physical presence takes Lee’s [9] definition of the sense of virtual objects acting as actual objects. In contrast, spatial presence takes Tamborini and Bowman’s definition [12], which refers to the sense of being located within a virtual space. Therefore, it can be concluded that presence refers to the perceived actuality of a virtual environment, including the perception of virtual physical and social interactions, virtual self, and virtual space.

Finally, cognitive absorption is a term conceptualised as a personality trait in Telogen's Multidimensional Personality Questionnaire [13]. In a review performed in 1990 of the then-current literature, Roche and McConkey identified the concept as "a readiness for experiences of deep involvement" and "an imperviousness to normally distracting events" [14]. Since then, some have argued that cognitive absorption is in some part equivalent to flow [15], however, some have argued they are separate concepts [1]. Throughout this project, a distinction is made between the two. Flow is defined as a state of deep engagement and concentration brought about by challenge. Absorption, however, is defined as a state of deep involvement wherein players dissociate from the environment around them, thus becoming less susceptible to distractions.

Within this report, immersion will constitute a mix of flow, presence, and cognitive absorption. Therefore, to enhance player immersion, one or multiple of these areas must be manipulated. Furthermore, to understand the complete effect of this project's final product on its users, all three of these metrics must be measured and evaluated.

2.1.2 Measuring Immersion

After discussing the exact meaning of immersion within this project, it is now possible to investigate methods for measuring immersion according to academic research. This section will explore evaluation methods for flow, presence, and cognitive absorption.

Among the various methodologies employed to measure player immersion, physiological techniques provide objective data. In their research, Drachen et al. found evidence that changes in heart rate and electrodermal activity can indicate varying levels of engagement and emotional response [16] in players. Eye-tracking technologies also provide insights by monitoring how players visually interact with game environments, revealing what attracts and holds their attention [17]. However, while these methods are unique and offer valuable quantitative data for measuring engagement, they do not provide specific measurements for the three underlying concepts of immersion. Therefore, qualitative feedback will be used to evaluate immersion in this project.

Qualitative measurement methods have been thoroughly investigated within academia for all three concepts. Flow is often measured using several well-established questionnaires. The Flow State Scale (FSS) [18], developed by Jackson and Marsh in 1996, was designed to assess flow during physical activities using a 5-point Likert scale, existing in an extended version of 36 items and short versions of 9 items. The Dispositional Flow Scale is another questionnaire designed to measure flow, created by Jackson et al. [19]. Jackson revisited the FSS and DFS in 2008, validating prior research and creating revised versions of the FSS-2 and DFS-2 [20]. Since the release of this paper, further research conducted by Hamari and Koivisto has found it successful in measuring the flow experience in the

gaming context [21]. Presence, the sense of 'being' in a virtual environment, also has a plethora of questionnaires by which it can be measured, the most prevalent of which is the Presence Questionnaire (PQ) developed by Witmer and Singer in 1998 [22]. This metric uses a 7-point Likert scale containing 19 items that gauge the extent to which users feel part of a virtual environment, perceive virtual objects as real, and experience virtual interactions as genuine. Methods for measuring cognitive absorption are less researched. However, one that is typically used is the Tellegen Absorption Scale (CAS), which measures the tendency to become mentally absorbed in everyday activities [23]. This scale, created by Tellegen and Atkinson, evaluates dimensions such as response to engaging stimuli, synaesthesia, and oblivious involvement in a 34-item, true or false questionnaire.

It is beneficial to use a combination of these questionnaires to capture the multifaceted nature of immersion. For example, the Game Experience Questionnaire created by Ijsselsteijn et al. measures dimensions of sensory and imaginative immersion and flow alongside others [24]. Another of these is the Game Engagement Questionnaire, created by Brockmyer et al., which measures flow, presence, and immersion in a single questionnaire [2]. These tools are designed to provide a holistic view of the gaming experience and have been validated in various studies. By utilising these validated questionnaires, it is possible to gather qualitative and quantitative data to understand and enhance player immersion effectively. These instruments offer insights into individual components of immersion and help create a nuanced understanding of the overall gaming experience.

2.2 State of the Art

2.2.1 Text-Based SoloRPGs

Traditional text-based solo RPGs offer a unique and immersive gaming experience. Players engage in storytelling, character development, and world-building on their own. These games have evolved from simple gamebooks to more complex systems incorporating various gameplay mechanics.

Figure 2: Choose Your Own Adventure Book

The origins of solo RPGs can be traced back to gamebooks, one of the first of which, called "The Roman Hat Mystery", was created in 1929 by Frederic Dannay [25]. Titles like the "Choose Your Own Adventure" series, seen in Figure 2, allowed players to make choices that determined the narrative's direction and launched SoloRPGs into popularity, providing an early form of interactive storytelling [26]. The "Fighting Fantasy" series by Steve

Jackson and Ian Livingstone added RPG elements, such as character creation and dice rolling, to the gamebook format, offering a more immersive solo RPG experience [27]. Other SoloRPGs included adventure modules designed for single-player experiences in addition to gamebooks. Games like "The Solo Dungeon" [28] for Dungeons & Dragons adapted multiplayer RPG mechanics for solo play. These modules provided structured adventures with predefined scenarios and challenges that players could tackle independently.

Comparatively, modern solo RPGs have diversified in terms of themes and mechanics. Currently, SoloRPGs often emphasise narrative and emotional engagement over complex mechanics. They encourage players to immerse themselves in the story, using evocative descriptions and personal reflection to drive the gameplay. For instance, games like "The Machine" focus on the character's psychological journey, while "Quill" challenges players to write persuasive letters within a fantasy setting.

Overall, traditional text-based solo RPGs have evolved to offer rich, varied experiences that cater to different tastes and preferences. They blend elements of storytelling, role-playing, and personal reflection, creating a deeply immersive and personal form of entertainment. Chapter 4 will further analyze the exact mechanics and variety of narrative styles in solo RPGs.

2.2.2 Large Language Models

LLMs have reached an advanced stage of development, marked by significant improvements in their capabilities and applications. These models exhibit a high degree of proficiency in understanding and generating human-like text, making them invaluable in various domains, including gaming, customer service, education, and content creation.

Large Language Models learn the nuances of language by learning patterns and structures from massive datasets, often containing billions of words. The training process involves feeding the model with text and adjusting its internal parameters based on how well it predicts the next word in a sentence. This is done using a " deep learning technique," which consists of many layers of artificial neurons that mimic the human brain. Each layer processes the text slightly differently, gradually building a sophisticated understanding of language. The model learns to recognise patterns, such as the sentence structure or the relationship between words, which it uses to generate coherent and contextually relevant responses.

Researchers continuously improve LLMs by increasing the amount of training data and refining the model architecture. They also use techniques to ensure the model's outputs are accurate and aligned with human values. For instance, the models are fine-tuned on specific tasks or datasets

to enhance their performance in particular areas. This iterative training and refining process results in increasingly proficient models at understanding and generating human language. In essence, LLMs are like highly knowledgeable and skilled language experts, created by exposing them to a vast amount of text and teaching them to understand and generate language through pattern recognition. The advancements in this field are a testament to the power of deep learning and the ongoing efforts of researchers to push the boundaries of what artificial intelligence can achieve.

For further details, one can refer to the following scientific papers: Brown et al.'s work on "Language Models are Few-Shot Learners" (2020) published in NeurIPS [1], which provides an in-depth look at the development and capabilities of large language models, and Vaswani et al.'s "Attention is All You Need" (2017) published in NIPS [2], which introduces the Transformer architecture, a foundational element in modern LLMs.

LLMs today can comprehend and produce text with remarkable fluency and coherence. They are trained on extensive datasets encompassing diverse linguistic patterns, cultural contexts, and specialised knowledge. This enables them to generate contextually relevant and nuanced responses, closely mimicking human conversation. The ability of these models to understand context, manage dialogue, and generate creative content makes them ideal for enhancing interactive experiences, such as SoloRPGs. One of the notable advancements in LLMs is their ability to personalise interactions. These models can adapt to individual user preferences and styles through techniques like few-shot learning and user-specific fine-tuning. This adaptability is crucial for applications like SoloRPGs, where personal immersion and a tailored narrative experience can significantly enhance engagement and satisfaction. Additionally, recent developments have extended the capabilities of LLMs beyond text. Modern models can process and generate content across multiple modalities, including text, images, and, in some cases, audio and video. This multimodal integration allows for more immersive and interactive applications, combining visual elements with textual narratives to enrich the user experience.

Today, LLMs are integral to a wide range of applications in natural language interfaces. They are used in chatbots and virtual assistants, such as OpenAI's ChatGPT and Google's Assistant, providing users with interactive and conversational experiences. These models facilitate customer support, automate routine inquiries, and enhance user engagement through personalised interactions. In content creation, LLMs assist in drafting articles, generating creative writing, and even composing poetry and music lyrics. They are also employed in coding assistance, where models like GitHub Copilot use LLMs to provide real-time code suggestions and documentation generation. Moreover, LLMs are increasingly utilised in specialised domains such as healthcare for medical record

summarisation, legal tech for document review and contract analysis, and finance for generating market reports and financial analysis. Their ability to process and synthesise large volumes of information makes them valuable tools in these industries.

In summary, LLMs have revolutionised the field of NLP by offering powerful tools for generating and understanding human language. Through continuous advancements and the development of robust evaluation metrics, these models are becoming more reliable and versatile, driving innovation across various sectors.

2.2.3 AI Dungeon

SoloRPGs that use LLMs in an attempt to create a new experience have already existed for some time. The most popular of these is *AI Dungeon,* which was first released in 2019 and is described on its website as "a text-based, AI-generated fantasy simulation with infinite possibilities" [29].

Figure 3: AI Dungeon Scenario Creation Screen

The first step in playing a game in AI Dungeon is choosing a world, story, and character. This can be done by finding community-made scenarios, creating custom scenarios by inputting a prompt, or selecting from in-built presets (see Figure 3). After entering a scenario, the game delivers a prompt for the player to respond to. For example, after entering a fantasy kingdom scenario roleplaying as a knight, the following prompt was delivered:

"You are Sir Knight, a knight living in the kingdom of Larion. You have a steel longsword and a wooden shield. You are on a quest to defeat the evil dragon of Larion. You've heard he lives up at the north of the kingdom. You set on the path to defeat him and walk into a dark forest. As you enter the forest you see a narrow, winding dirt path leading deeper into the dense foliage. The trees tower above you, their branches intertwining to form a natural canopy that filters the sunlight. A cool breeze rustles the leaves, casting dappled shadows on the ground. The air is thick with the scent of damp earth and moss.

As you proceed, the forest becomes quieter."

After receiving the prompt, a player can choose to either "take a turn" by providing text-based input, continue the generation of the story without input, ask the AI to regenerate its previous output or undo their previous turn. If the player chooses to take a turn, they have four ways in which they can provide input. They can either describe an action their character will take, something their character will say, something they see, or provide direct input into what happens next in the story.

AI Dungeon succeeds in providing a SoloRPG-like experience by allowing players to explore a fantasy world with freedom of both narrative and choice. However, while using the program several pitfalls were found that demonstrated AI Dungeon, while innovative in its approach, falls short in several key areas compared to typical SoloRPGs. While SoloRPGs will direct the flow of play by selectively offering meticulously crafted experiences with detailed rule sets, structured storylines, or other defined game mechanics, AI Dungeon offers none. These elements provide a consistent and immersive experience that AI Dungeon, without such elements, struggles to replicate. The AI-generated content often lacks depth and coherence due to its unrestrained nature, resulting in disjointed and sometimes frustrating experiences for the player.

Firstly, AI Dungeon's implementation of AI-generated text can lead to significant illogical plot developments. In traditional SoloRPGs with defined storylines, the narrative is carefully curated to ensure continuity and logical progression, allowing players to immerse themselves in the game's world fully. However, AI Dungeon's text generation can produce random and unpredictable outputs that often break immersion and detract from the overall storytelling experience. This unpredictability usually means that players must frequently reset or redirect the narrative, interrupting the game's flow and diminishing overall enjoyment.

Furthermore, the lack of defined mechanics in AI Dungeon allows players to exploit the game, undermining its integrity and immersion. Suspension of disbelief, as defined by Oxford Languages [30], is the "willingness to suspend one's critical faculties and believe the unbelievable."; sacrifice of realism and logic for the sake of enjoyment. As players move through the game, they may make any decision they desire or instruct the story to go in each direction. The AI must then find a way to work this into the story, no matter how absurd or unrelated to the current narrative a direction is. This leads to scenarios that feel artificial, unimpactful and unrewarding. For example, if during the game a monster appears, the player may simply give the instruction, "I kill it", and the AI will have no choice but to create a contrived explanation on behalf of the player. As a result, the gaming experience becomes less about engaging with a rich narrative and more about exploiting an unpredictable AI, ultimately reducing overall enjoyment and diminishing immersion.

Finally, the reliance on AI technology in AI Dungeon can lead to repetitive and unoriginal content. Traditional SoloRPGs often feature diverse and imaginative storylines crafted by experienced game designers, ensuring a rich and varied gameplay experience. In contrast, AI Dungeon's AI-generated content can become monotonous over time, with the AI frequently recycling phrases, plot elements, and character archetypes. This repetition can quickly diminish the game's appeal, lacking the creativity and originality found in traditional SoloRPGs.

In summary, while AI Dungeon offers a novel approach to solo role-playing, it pales in comparison to the coherence, skill, and depth available in traditional SoloRPGs. Its reliance on AI-generated content results in frustrating and illogical story developments, a lack of skill required for the player, and repetitive storytelling that fails to match the rich narratives of its traditional counterparts.

2.2.4 Rabbit R1 and Humane AI Pin

Understanding how a natural language interface can facilitate interaction between humans and AI is crucial to the success of this project. Therefore, two AI-enhanced devices that go beyond simple text input have been investigated. The Rabbit R1 (see Figure 5) and the Humane AI Pin (see Figure 4) are

Figure 5: R1 Rabbit

Figure 4: Humane AI Pin

both portable devices that facilitate tangible interaction between humans and AI. While both attempt to serve the same purpose as AI assistants, they each offer a unique approach.

The R1, designed by Rabbit, utilises what Rabbit calls a Large Action Model (LAM), which claims to be a “new type of foundation model that understands human intentions on computers” [31]. The LAM essentially attempts to act as a universal controller for applications. In their presentation at CES 2024, Rabbit stated that this was achieved by training the LAM on humans interacting with apps like Spotify and Uber [32]. The LAM runs on rabbit-os, the operating system of the R1. Users can interact with the R1 by either using the scroll wheel, pressing the side button, speaking to it, or taking pictures using the swivelling camera. Furthermore, Rabbit has stated that if a user wishes to teach the R1 a custom application or task such as Photoshop, they are able to train the LAM themselves [32].

The AI Pin, designed by Humane, is a wearable device that attempts to act as a digital assistant using their CosmOS operating system [33]. Users can interact with the AI Pin by using the touchpad, speaking to it, and using its Laser Ink Display system pictured in Figure 6. Humane states that the Pin can then perform many tasks, such as updating the user on news or sending messages. Additionally, by holding objects in front of you and asking the pin questions, it will use its camera to see what you're carrying and respond accordingly [33].

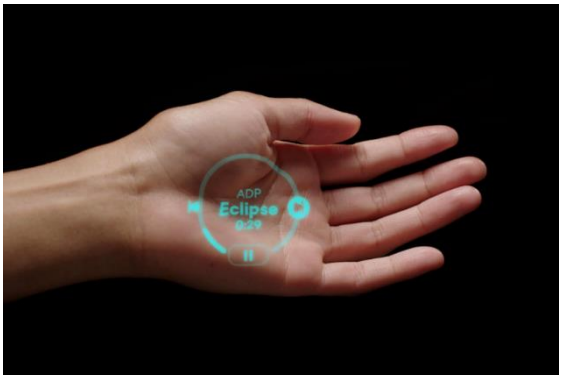

Figure 6: Laser Ink Display

Public perception of both devices has been largely negative, with many critics arguing that their functionalities could be achieved through more straightforward smartphone applications [34]. The Rabbit R1, in particular, has faced severe scrutiny after investigative reports from the YouTuber Coffeezilla revealed that its touted Large Action Model was not a novel AI system but rather a repackaged version of existing technologies like ChatGPT, supplemented with a program called Playwright, a web automation tool [35]. Furthermore, early reviews describe the R1 as a mess, citing numerous non-functional features and an overall sense that the product was rushed to market. Similarly, the Humane AI Pin has been criticised for lacking essential features and failing to meet its ambitious goals, leading to scepticism about its practicality and market readiness. Prominent tech commentators like Marques Brownlee have criticised the trend of releasing unfinished products, arguing that both the R1 and AI Pin exemplify this problematic approach [36].

2.3 Conclusion State of the Art and Literature Review

This literature review has explored the multifaceted concept of immersion within gaming, dissecting it into the components of flow, presence, and cognitive absorption. Flow, as defined by Csikszentmihalyi, represents a state of deep engagement and enjoyment when player skill matches game challenges, leading to an immersive experience. According to Witmer and Singer, presence encompasses the sensation of being part of the virtual environment, further categorised into self-presence, social presence, and physical and spatial presence. Cognitive absorption, a concept linked to personality traits, involves a deep involvement that diminishes awareness of the physical environment, making players less susceptible to distractions.

Measuring these components of immersion can be achieved through both physiological and qualitative methods. Physiological techniques, such as monitoring heart rate and electrodermal activity, provide objective data on player engagement. However, qualitative feedback is the most

well-researched and validated tool for a comprehensive understanding of immersion, as it directly addresses players' subjective experiences. An accurate evaluation of a player's overall immersion may be found by measuring each flow, presence, and cognitive absorption.

The review also delved into the evolution of text-based SoloRPGs and the impact of large language models (LLMs) on interactive storytelling. Traditional SoloRPGs offer structured, immersive experiences through meticulously crafted narratives and game mechanics. Modern adaptations like AI Dungeon attempt to leverage LLMs for dynamic storytelling but often fall short in coherence, engagement, and depth compared to their traditional counterparts.

Examining AI-enhanced devices like the Rabbit R1 and Humane AI Pin highlighted the challenges in integrating advanced AI into practical, consumer-friendly products. Despite their innovative approaches, both devices have faced criticism for their lack of functionality and the perception that more straightforward smartphone applications could replicate their capabilities. Investigations revealed shortcomings in the underlying technology, casting doubt on their readiness for the market.

In conclusion, this project will require a careful balance of the elements of traditional SoloRPGs, precise implementation of LLM technology, and comprehensive and research-supported methods for measuring immersion. While current technologies show promise, they also highlight the ongoing challenges in delivering fully immersive and functional AI-driven experiences. Lastly, throughout the conducted research, no results were found on measuring how immersion can be influenced by the inclusion of large language models into existing activities. Thus, this project aims to bridge this hole.

Chapter 3. Methods and Techniques

In this chapter, the techniques and methods used in the completion of this project will be explained. In the first section, the creative technology design process will be described, and its implementation within this project will be discussed.

3.1 The Creative Technology Design Process

Mader and Eggink outline the design process used for creative technology projects in their paper "A Design Process for Creative Technology." [37]. A diagram of this process can be seen in Figure 7. Note that each of the latter stages has an arrow redirecting to its former stage. This indicates that it is possible to return to a previous stage at any point. This process consists of four phases to produce a carefully thought-out product: Ideation, Specification, Realization, and Evaluation. The first stage is ideation, wherein either technology, a creative idea, or user needs/stakeholder requirements are co-examined to create ideas. Ideation is followed by Specification. In this stage, experience specifications, functional specifications, and early prototypes are created to refine requirements and explore various designs. This process is conducted in a "short evaluation and feedback loop" [37] in order to quickly and effectively determine design specifications. In the third stage, Realisation, the design requirements are used to construct a product. Following this, the Evaluation stage is conducted. In this stage, products are evaluated through user testing and function testing, and if the result is unsatisfactory, the product should be returned to a previous stage and reconsidered.

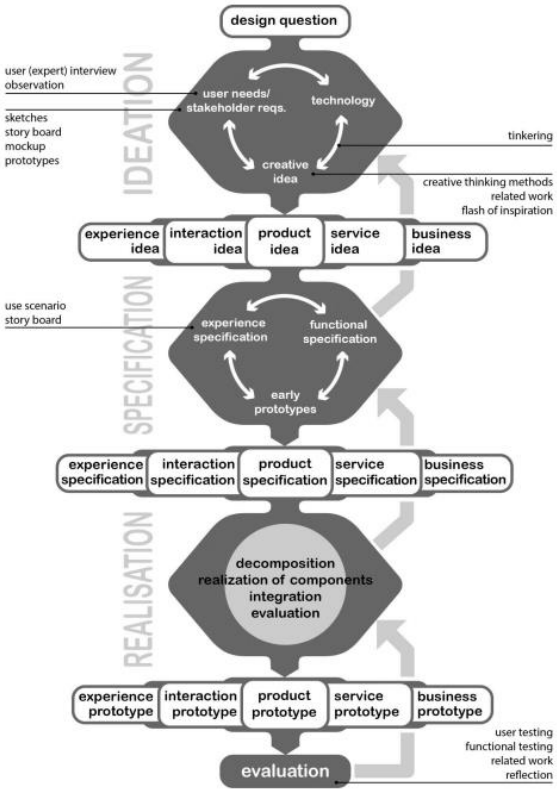

Figure 7: Diagram of the Creative Technology Design Process

3.2 Stakeholder Analysis Method

Stakeholders of this project are persons, groups, or organisations with a vested interest in the outcome of this project. Therefore, an analysis of their various roles, importance, and interests is considered is a vital step in producing a successful project. Stakeholders can be multiple people or organisations, such as the end-users, in this case, the players, investors, etc. To conduct an analysis of this project's stakeholders, they must first be identified and their roles described. A table will be made using the template shown in Table 1 to achieve this. In the first column, the name of the

stakeholder is given, and in the second, a short description of their role within the context of this project is described. As the eventual analysis will be conducted using the power versus interest grid designed by Mendolow [38], the power and interest of each stakeholder will be explained in the third and fourth columns. In this method, power refers to the ability of a stakeholder to influence decisions within the project, and interest refers to the extent to which a stakeholder is vested in the project.

Stakeholder	**Role**	**Interest**	**Power**
X	X	X	X

Table 1: Stakeholder Role/Interest/Power Template

After identification, the analysis can be conducted using the power/interest grid shown in Figure 8. As can be seen, interest, plotted on the X-axis, is compared to power on the Y-axis. Therefore, the more interest a stakeholder has in the project, the further along they will be marked on the graph, and the more interest they have, the higher. The 4 different regions - Monitor, Keep Satisfied, Keep Informed, and Manage Closely - that a stakeholder might fall in determine their importance and the manner in which they should be involved in the project. For example, stakeholders that are plotted low and close to the origin hold little influence over the project and are not greatly interested, meaning they must only be monitored.

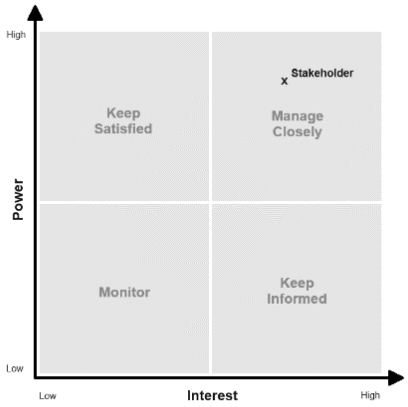

Figure 8: Power/Interest Grid Template

3.3 SoloRPG Mechanics Analysis Method

In order to understand what gameplay elements and mechanics are essential to a SoloRPG, an analysis will be conducted on a selection of SoloRPGs. This analysis will be performed in three parts. In the first, each game will be played in totality, and notes will be taken on game mechanics, storytelling elements, and other observations during gameplay. Next, the notes will be combined and labelled in order to create a list of possible game mechanics and scaled metrics for other game elements. The template for these can be found in Table 2 and Table 3. Finally, each game will be replayed and evaluated using the conceived metrics. In order to ensure a variety of SoloRPGs are examined, the games that will be analysed will be compiled from Dicebreaker's "5 best solo RPGs you can play by yourself", Paul Joyce's "Five Best Solo RPG Games to try in 2024", and WarGamer's "Best

solo RPGs 2024" lists. This should also ensure the preferences of any one individual or company do not dictate the findings of this analysis.

Game	Mechanic 1	Mechanic 2	Mechanic 3	Mechanic 4
X		X		X

Table 2: SoloRPG Mechanic Analysis Template

Game	Element 1				
	1	2	3	4	5
X				X	

Table 3: SoloRPG Element Analysis Template

3.4 Requirement Identification and Categorisation Method

In order to give direction to this project, the requirements of the product have to be evaluated. The first step is to identify these requirements. These requirements will be informed by consulting a mixture of sources. A stakeholder analysis will be conducted in order to determine the key groups that need to be satisfied. Furthermore, an analysis of current text-based SoloRPGs will be undertaken in order to identify the key elements that must be captured and prompt the direction of the game.

Once the requirements have been identified, they can be categorised using Clegg and Barker's MoSCoW method [39]. This method prioritises the requirements by sorting them into four groups: Must Have, Should Have, Could Have, and Will Not Have. This method sets comprehensive guidelines that can direct the project towards success. The template for this method can be found in Table 4.

Requirement	Must Have	Should Have	Could Have	Will Not Have
X		X		

Table 4: MoSCoW Template

3.5 Evaluation Method

To investigate the research question, an AB test will be performed. This evaluation will compare a modified version of a SoloRPG that has been enhanced using a large language model to its traditional version.

Firstly, players will be asked to read and fill out an information sheet and consent form that relays the details of the evaluation and confirms they understand their rights and responsibilities. Next, players will be told which version of the game they will play first, which will be decided by a random number generator. The players will then be asked to play 15-20 minutes of the given version, after which they will be asked to fill out a questionnaire and then repeat the process for the other

version. To gather users, students from the University of Twente were invited to join, and invitations were sent out on public forums in communities that centre around SoloRPGs, such as the 'Lone Wolf Roleplaying' Discord Server.

The Game Engagement Questionnaire (GEQ), developed by Brockmyer et al., will be used for this evaluation. [2]. The GEQ is a 19-item questionnaire where participants answer the given questions, which can be found in Appendix A, on a three-point Likert scale; however, a five-point scale will be used in this evaluation instead. This questionnaire is ideal for this evaluation as it is short and easy for participants to understand. Furthermore, each question in the GEQ is classified into a construct of either flow, presence, or immersion, which follows this paper's definition of immersion.

The process of analysing the collected data starts by assigning a value to each point on the Likert scale (no = 0, sort of = 2, yes = 4) and summing the total of respondents' answers to each construct. This yields a score for flow, presence, and absorption per user for each version of the game. Next, all the scores for each concept and version will be compared, and a two-tailed, paired T-test will be performed for each concept to determine if there is a significant difference in the mean scores for each version. For example, to perform the t-test on the flow category, a null hypothesis (H_0) is given: There is no difference between the traditional and modified versions in terms of flow. An alternative hypothesis (H_1) is given as follows: There is a difference between the modified and traditional versions in terms of flow. Next the test statistic is calculated using Equation 1.

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

Equation 1: t-test equation

Where $\bar{d}$ is the mean of the differences between paired evaluations, s is the standard deviation of the differences, and n is the number of pairs. Finally, compare the t-test to the critical value, and either reject or fail to reject the null hypothesis.

Chapter 4. Ideation

4.1 Stakeholder Analysis

The stakeholders were first identified in Table 5.

Stakeholder	Role	Interest	Power
Players	CUSTOMER The players will ultimately use the product when they engage in a SoloRPG.	VERY HIGH Players hold the most interest as the users of the product.	HIGH As players are the customers, the project aims to meet their requirements and satisfy their needs.
Generative AI Corporations	SUPPLIER Companies such as OpenAI provide LLMs that can be used for the product. They host these models on servers that can be accessed through APIs.	MEDIUM It is possible that companies that provide LLMs are interested in their use cases.	MEDIUM If a company decides this project cannot make use of its model, a large part of the project is removed. However, as there are many providers of LLMs, a switch can be made somewhat easily.
Open source/Library Developers	SUPPLIER Specific libraries will be used throughout this project that have been published and contributed to by many developers.	LOW Most libraries are made for general use, and the developers are unlikely to be interested in the products they are used in.	LOW It is difficult to prohibit this project from using open-source software. Therefore, these stakeholders have little power.
Game Developers	PARTNER/CUSTOMER Game developers may wish to use the technology involved in this project to create games and may use the product themselves as a reference.	HIGH Game developers may be highly interested in the project as a product they can incorporate or reference in their own games.	MEDIUM The product is not designed for game developers, although their wants may be of interest to the project.

Table 5: Stakeholder Role/Power/Interest Analysis

After identifying the stakeholders and their interest/power levels, they were placed in the Power versus Interest Grid shown in Figure 9.

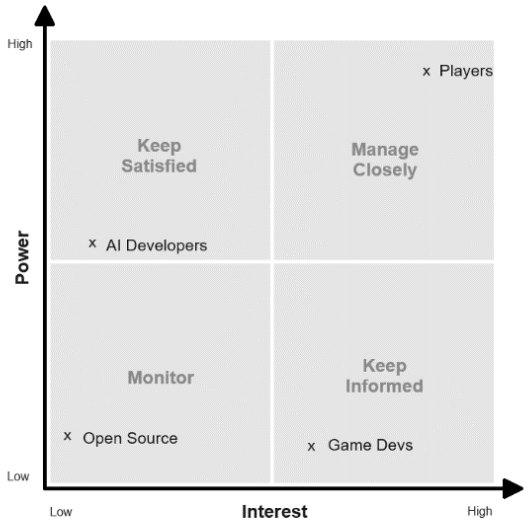

Figure 9: Stakeholder Power/Interest Grid

As can be seen from this analysis, the players both hold high power and interest in this project. Therefore, they fall into the manage closely classification, meaning that for this project to be successful players should be consulted frequently through development. AI developers and game developers fall on opposite sides of the grid. While AI developers have little interest in this project, it is still vital that their models are not misused, and thus, they need to be kept satisfied by abiding by the regulations they set. Contrastingly, game developers hold little power over the project, as the product is not being designed with them in mind. However, they may hold interest in the outcome as they wish to use similar technology in their own projects. Therefore, game developers who express interest in the project should be kept informed. Finally, open source developers have low interest and power within the project and thus need only be monitored.

4.2 SoloRPG Analysis

In the first step of this analysis, one entire run of a selection of SoloRPGs was conducted, and notes were taken for each game. These notes contain mechanics present within the played game and noted elements that were perceived whilst playing.

In the second step of the analysis, the notes were compiled and labelled to determine various game mechanics present throughout the games. The list of identified game mechanics can be found in Table 6.

Gameplay Mechanic	Description
Narrative Focus	If a game was mainly driven or had a significant focus on the narrative.
Randomness	If a game uses random elements (dice, cards, or both) to generate events, encounters, or environments.
Prompts	If a game provides narrative prompts to guide the story and offer direction
Lookup Tables	If a game includes tables to add structure and provide randomised outcomes or events.
Journaling	If a game involves writing as a core component, either through letters, diaries, or logs.
Exploration and Discovery	If a game asks players to explore unknown settings and uncover secrets.
Resource Management	If a game includes managing resources or characters, adding a strategic element.
Physical Components	If a game uses physical components (excluding dice and cards) like a player marker, a board, or block towers to influence gameplay.
Survival Elements	If a game contained failure-like conditions that involve a player's health.
Character Development	If a game involves the development of characters or entities over time, reflecting changes and growth.
NPC Interaction	If a game involves interactions with non-player characters, which can influence the story or game outcomes

Table 6: SoloRPG Mechanic Descriptions

Additionally, notes on the various narrative and storytelling elements were used to create scaled metrics that will be used in the third step. The list of these elements can be found in Table 7.

Element	Description
Freedom of Decision	The degree to which players can make choices that significantly impact the game's direction and outcome.
Narrative Freedom	The extent to which players are left to narrate the story on their own, without predefined prompts.
Mechanical Freedom	The flexibility players have in using game mechanics and systems to shape their experience.
Environmental Fidelity	The ability to explore and interact with the game world in varied and unrestricted ways.
Story Emphasis	The importance placed on narrative and storytelling within the game.
Complexity	The level of complexity within the game, such as the amount of mechanics, rules or pieces that may make the game more difficult to play

Table 7: SoloRPG Element Descriptions

Finally, all games were replayed, compared to the list of mechanics, and rated for each of the scaled metrics. The results can be seen in Tables 8 and 9.

Game	Narrative Focus	Random-ness	Prompts	Tables	Journaling	Exploration & Discovery	Resource Management	Physical Components	Survival Elements	Character Development	NPC Interaction
Gentleman Bandit	X	X (Cards)	X								X
Ex Novo	X	X (Dice)		X							
The Machine	X	X (Cards)	X	X		X					
Alone Among the Stars	X	X (Cards)	X			X					x
Quill	X	X (Dice)	x		X						X
Four Against Darkness		X (Dice)		X		X	X			X	
Alone Against the Frost	X	X (Dice)	X				X		X	X	X
Ironsworn	X	X (Dice)	X			X	X		X	X	X
The Wretched	X	X (Cards)	X					X	X		
2D6 Dungeon		X (Dice)		x		X	X				
Thousand Year Old Vampire	X	X (Dice)	X	X	X		x			X	X
Be Like a Crow	X	X (Cards)	X		X						

Table 8: SoloRPG Mechanic Analysis

Game	Freedom of Decision	Narrative Freedom	Mechanical Freedom	Environmental Fidelity	Story Emphasis	Complexity
Gentleman Bandit	Medium	High	Low	Low	Medium	Low
Ex Novo	High	Medium	Medium	High	Medium	Medium
The Machine	Medium	High	Low	Medium	High	Medium
Alone Among the Stars	Medium	High	Low	High	High	High
Quill	Medium	High	Low	Low	High	Low
Four Against Darkness	Medium	Medium	High	Medium	High	Medium
Alone Against the Frost	Medium	High	Medium	Medium	High	High
Ironsworn	High	High	High	High	High	High
The Wretched	High	High	Medium	Medium	High	Medium
2D6 Dungeon	Low	Low	High	Low	Medium	Low
Thousand Year Old Vampire	High	High	Medium	Medium	High	Medium
Be Like a Crow	Medium	High	Low	High	High	Low

Table 9: SoloRPG Element Analysis

As can be seen from the mechanical analysis, almost all games have a narrative focus and use randomness as a mechanic. Furthermore, every single game uses either prompts or tables in order to create situations, conflicts, or other content for the player. This is a significant finding as it clearly demonstrates how the majority of SoloRPGs share these core mechanics. Beyond these, the following three most widely adopted mechanics are exploration & discovery, resource management, and NPC interaction. However, each of these mechanics was present in no more than six out of the twelve games. This demonstrates that while the majority of SoloRPGs adopt similar core mechanics, there is also a lot of mechanical variety within the genre.

Additionally, it can be seen from the element analysis that story emphasis is ranked high in nearly every game, with no game scoring low. Freedom of decision and freedom of narrative are also scored high on average. The other elements that were considered can be seen to vary largely across the different games. This exemplifies that SoloRPGs, at their core, are games about narrative and take advantage of the ability to provide players with freedom. Furthermore, it indicates that while narrative and freedom are shared amongst most SoloRPGs, other elements differ between games and thus make SoloRPGs a diverse field.

4.3 Preliminary Requirements

After conducting the stakeholder analysis and SoloRPG analysis, the findings as well as previous research was used to create preliminary requirements that would guide the initial concept development and selection. These requirements can be seen in Table 10.

Requirement	Must	Should	Could	Won't
Target at least one form of immersion	X			
Target at least two forms of immersion		X		
Target all three forms of immersion			X	
Be able to adapt from an existing SoloRPG	X			
Natural Language interface integration using a LLM	X			
Include randomness as a mechanic	X			
Allow high freedom of player decisions	X			
Emphasis on story	X			
Use lookup tables for random events, prompts, etc.	X			
Allow player exploration		X		
NPC interaction		X		
Include resource management		X		
User-friendly interface		X		
Include survival elements			X	
Include Journaling Capability			X	
Voice Input/Output			X	
Physical Components			X	
Multi-Platform Support				X
Support Multiple Languages				X

Table 10: Preliminary Requirements

Firstly, the way in which immersion would be targeted was considered. As the three concepts, flow, presence, and absorption, were found to be all factors of immersion, it was determined that by targeting and enhancing one or multiple of these, immersion could be increased. Therefore, at least one of these factors must be targeted, and additional ones should or could be. Additionally, the product must be able to adapt and modify the experience of a traditional SoloRPG game. This will allow for user testing to directly compare a modified version of a SoloRPG with a conventional version and should enable quantitative data to be collected. Furthermore, this project aims to investigate whether LLMs can be used to enhance immersion. Therefore, the following requirement that must be met is a large language model being used to, in some way, enhance the game through a natural language interface. Furthermore, the mechanics and elements found in almost every SoloRPG must be included in the product. This includes narrative freedom, randomness, prompts or tables, and an emphasis on the story.

Next, at least one of the three mechanics that were present in half of the games should be included. In doing this, the product will retain the core experience of a SoloRPG but will also have a personality of its own. Furthermore, it allows this project to explore which other mechanics lend themselves to LLM integration. Next, the user interface should be friendly and easy to use. This is in order to minimise player confusion, which could affect immersion and, thus, muddle the evaluation of the research question.

The other mechanics found in a few games could be in the product. However, they are not vital as they represent minorities of the SoloRPG genre. Considering that a natural language interface will be used in the project, voice input and output could be used. However, this is not necessary, as typing on a keyboard is also an acceptable form of communication. The same can be said for physical components, as digital buttons and other elements are also interfaces.

Finally, this project will not attempt to maximise compatibility between multiple operating systems, as although it may reduce the number of available testers, it is a large task that does not affect the results of the actual evaluations. Furthermore, multiple languages will not be supported for the same reason.

4.4 Preliminary Concepts

4.4.1 NPC Box

In this concept, player interaction with Non-Player Characters (NPCs) is conducted through a natural language interface powered by an LLM. NPC interaction within is largely non-present in SoloRPGs or is left for the player to imagine as it is infeasible to create dialogue for every possible scenario a player may encounter. Furthermore, prewritten dialogue within the game limits a player's options and freedom as they must select from prewritten responses. This concept allows players to interact with characters from the game world in two steps. The NPC Box concept changes this aspect. In order to interact with an NPC, the player must insert the corresponding character card into the box. The box is then able to read the card and develop the character's personality traits, memories, and other information, such as the character's voice. Finally, the player speaks to the box and has a conversation with the character as if they were in the room.

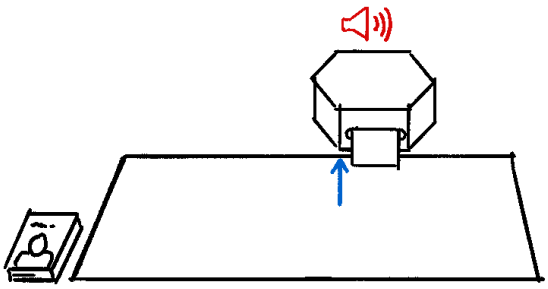

Figure 10: NPC Box

4.4.2 Auto Board

The Auto Board removes the responsibility of storytelling from the player. SoloRPGs often provide gameplay mechanics and general information on a game's story elements, such as the environment. However, they cannot provide a full narrative experience, as each game holds many possibilities. Therefore, the players are responsible for imagining and narrating the story for themselves, which may be difficult for certain players and damage immersion. The Auto Board changes this by automatically detecting the state of the game each time the player makes a move and uses an LLM to narrate the story automatically. This removes unnecessary strain from the player and enables them to immerse themselves more deeply into the game.

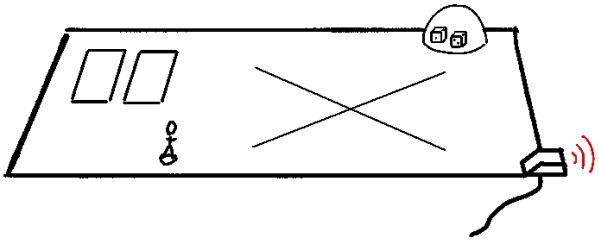

Figure 11: Auto Board

4.4.3 Handy RPG

The Handy RPG is a Handheld device that uses an LLM and Natural Language Interface to create a SoloRPG experience. This concept automates every aspect of a SoloRPG but retains the player's ability to make decisions. The device accepts a drive that has a preloaded game on it, including rules, characters, story, and other gameplay mechanics. The Handy then runs the game and uses the built-in LLM to narrate the story. Whenever it is time for the player to make a move, they are still able to make whatever decision they would like, and the Handy will automatically perform any additional actions such as rolling die or looking up values. This removes many actions that inhibit a player's immersion while retaining their freedom.

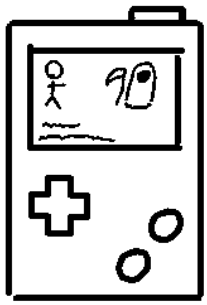

Figure 12: Handy RPG

4.4.4 Desktop RPG

Desktop RPG is a video game that emulates the experience of a SoloRPG while enhancing it using a LLM. In this concept, a player can engage in a SoloRPG from their computer by dragging and dropping items as they would on a physical board. In addition to this, players can input moves or other text-based commands via text input. A LLM is also included in this application, which can narrate the

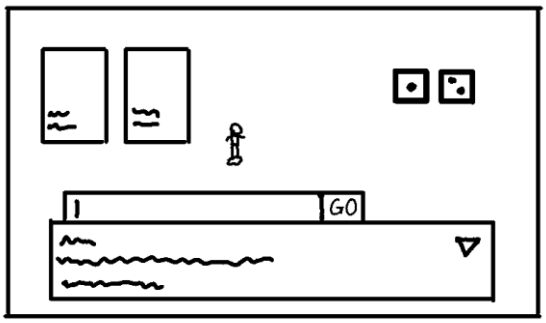

Figure 13: Desktop RPG

story, act as NPCs, and respond to player input. Players are able to enjoy the essential SoloRPG experience while also having the freedom to act however they want.

4.5 Concept Selection

After creating the initial concepts, one must be chosen to investigate further. Each idea was discussed with the client/player. This section describes the arguments for and against each concept, and a final choice is made on which to pursue.

The NPC Box presents a unique idea among the four concepts, focusing solely on enhancing NPC interaction. This concept would most likely be apt for enhancing player presence. As social presence is the degree to which social interaction feels real, by having the player able to speak to an NPC actively, their perception of the realness of their conversation could increase. However, from the background research that was conducted, it was found that stand-alone devices such as the r1 rabbit that use voice as a communication method do not seem ready for market use currently. This indicates that as a sole developer, it would be wise to steer clear of this concept.

The Auto Board also has its own unique elements, as it is the only concept that uses physical pieces. Alongside this, it is able to narrate for the player. This means that this concept could enhance both presence, by describing and amplifying the perceived realness of the world, and absorption, as physical pieces could make the gameplay feel intuitive. However, there are some significant concerns with this idea. It would require a high level of engineering, electronics, and software to work, which lies outside the scope of this project. Additionally, it would require that it be adapted from a SoloRPG that uses physical pieces, which heavily narrows down the choice of games. This is not justifiable, as physical pieces were found in the analysis not to be an essential element of SoloRPGs.

The Handy RPG would allow for a lot of freedom, as running the game entirely on a device where all of the code is written by the researcher allows for natural language interfaces to be implemented wherever possible. However, the Handy RPG is also a standalone device which, as previously mentioned, indicates that it may be unwise to pursue. Furthermore, the complexity of creating an entire console may lead to the actual SoloRPG experience being diminished. The Desktop RPG, however, has the same benefits as the Handy RPG, but does share its drawbacks. The Desktop RPG also has the additional functionality of being able to be shared easily online as an application.

Therefore, after discussing the benefits and drawbacks of each concept, a choice can now be made on which should be further investigated. For this project, the Desktop RPG was selected. This was primarily because it allows multiple facets of narration to be integrated with an LLM. Narration, random prompts, and NPC interactions are all features that can clearly be modified using AI in this

concept. Additionally, by having the entire game state in a single application, the player's current position and progress are always known. Thus, the program can easily decide what it should be doing at any given time, and the implementation of natural language interfaces will be simple.

Chapter 5. Specification

In the specification phase, the final concept is further specified in order to facilitate the creation of the end product. Firstly, a series of personas are created in order to conceptualise how various users may interact with the product. Next, a collection of lo-fi prototypes based on a selection of games from the prior analysis will be created and evaluated with the client. A final selection of which game the product is adapting will then be made. Lastly, the final design requirement will be presented.

5.1 Personas

5.2.1 Persona 1 – Olivia

Olivia is a 34-year-old manager at a tech company, balancing her demanding job with family life. Despite her busy schedule, Olivia actively seeks out ways to relax and unwind during her limited free time. Recently, she discovered solo RPGs through an online forum and has become intrigued by the concept. Although she has tried a few beginner-friendly rulebooks, she often finds the setup and rules complicated and time-consuming to learn, which can be frustrating given her limited downtime.

Given her tech-savvy background, Olivia is comfortable working with computers and prefers a digital solution for her gaming needs. She enjoys the freedom of making her own decisions within the game, finding it a refreshing escape from her structured work environment. Olivia primarily engages with SoloRPGs during her daily commute on the train to and from work as she doesn't typically use the app at home, reserving that time for family activities and other responsibilities. As her train ride is only 15 minutes, Olivia requires the games she plays to be short or if they are longer, have a save function. Thus, Olivia needs a solo RPG app that is user-friendly, quick to set up, easily accessible on her laptop and accommodates her short playtime, making her commute more enjoyable and stress-free.

5.2.2 Persona 2 – Alex

Alex is a 20-year-old computer science major with a long-standing passion for tabletop RPGs since he was a high schooler. Living in a shared dormitory, Alex often finds it difficult to coordinate game sessions with friends due to his demanding academic schedule. This has led him to explore the world of SoloRPGs as a way to continue his hobby on his own terms.

Given the shared living space, Alex prefers typing over speaking to ensure he doesn't disturb his roommates. He has tried digital SoloRPG experiences but misses the tactile sensation of rolling dice, a cherished aspect of traditional tabletop games, and seeks a digital solution that can replicate this.

The dorm environment can be loud, so Alex appreciates having background music within the app to help him focus and immerse himself in the game.

As a tech-savvy individual, Alex is comfortable navigating complex systems but has little patience for applications that lag or take too long to process commands. As a student, he also has only a small amount of storage space left on his laptop and, therefore, needs the games he downloads to be relatively small. He enjoys engaging and challenging game mechanics, finding satisfaction in intricate gameplay and strategic thinking. For Alex, a SoloRPG app needs to offer a seamless, responsive experience with various immersive features, himself, to keep him engaged despite the distractions of dorm life.

5.2.3 Persona 3 – Margaret

Margaret is a 68-year-old retiree living alone with a lifelong passion for storytelling. Over the years, she has indulged in reading novels and writing short stories, finding joy in the creation and exploration of narratives. With lots of free time at home, Margaret has recently become interested in SoloRPGs as a new way to engage with her love for storytelling.

Due to her preference for staying at home, Margaret seeks a SoloRPG app that she can comfortably use in her familiar environment. As she finds typing to be slow and cumbersome, she would greatly prefer an app that allows for speech input, making the experience more seamless and enjoyable for her. Not being particularly tech-savvy, Margaret needs a very user-friendly interface that is easy to navigate and understand without requiring a steep learning curve.

Her primary joy in SoloRPGs comes from the storytelling aspect, where she can immerse herself in rich narratives and make decisions that shape the plot. For Margaret, the ideal solo RPG app would be one that is accessible, voice-operated, and designed with a simple, intuitive interface, allowing her to fully engage with her beloved pastime of storytelling.

5.2.4 Conclusion

From the detailed analysis of Olivia, Alex, and Margaret, it is evident that the requirements for a solo RPG app must cater to a diverse range of user needs and preferences. Each persona highlights specific features and functionalities that would enhance their gaming experience, emphasising the need for a versatile and adaptive application. The immediate considerations from this analysis include a user-friendly interface, offline capabilities, speech and typing input options, background music, dice rolling, file size, and engaging narrative-driven content. However, not all of these features are possible, and it is therefore vital to consider and determine which are most crucial to the success of this project.

5.2 Lo-Fi Prototypes

5.2.1 Game Selection

For the Lo-Fi prototype phase, three of the previously analysed SoloRPGs were chosen to evaluate. In order to select these, each of the games was compared against selection criteria.

Firstly, as the element analysis found that narrative freedom and narrative emphasis were core features of SoloRPGs, only games that scored at least one high and one medium in these metrics were considered, eliminating three games from the possible pool. Next, games that were found to rank high on complexity were removed, taking an additional three games out of the possible candidates. This is both so that players can easily understand the game within a short time frame and to ease the amount of software development that must be done.

Next, if the mechanical analysis found that a game did not contain any of the three elements found in half of the games, it would be removed. This was done so that the final selection would contain games with a variety of mechanics, allowing for the most to be learnt from the evaluation. This left four games, of which Quill, Four Against Darkness, and Thousand Year Old Vampire were selected.

5.2.2 Prototype Design

For both of the selected games, a custom version of ChatGPT was created using OpenAI's GPT functionality. Using this system, a chatbot was created for both Quill and Thousand Year Old Vampire by feeding each the full game instructions in a text file. Additionally, each bot was given specific instructions for how they would act and respond to the player. This allowed for the basic mechanics of the game to be played and experimented with without a user interface. Furthermore, it was possible to investigate possible pitfalls of the GPT models and possible ways to negate these limitations.

Quill

Quill is a game that is played by selecting a character and scenario and then writing a 5-paragraph letter to the subject of the selected scenario. Each paragraph must contain one word from the scenario's 'inkpot', and a score is calculated throughout gameplay by rolling some dice. Finally, after completing every paragraph, the player must imagine the response they receive from the subject they wrote to, and this response varies according to their acquired score. For this prototype, the GPT was given the following instructions:

You will act as the narrator of a game of quill. The rulebook for quill is a part of your knowledge. Present the player with the decisions they have to make. When cards must be drawn or dice have to be rolled, ask the player to do this themselves. When the player gives you an answer, you may not change it, but you can elaborate on the story and narrate the game further. Do not reveal the consequences of different scenarios. Do not explain the rules as the game is played.

After presenting the scenario, say nothing more. The player will reveal their score to you after writing all 5 paragraphs.

Within this prototype, the job that was given to the GPT was to first navigate the player through the character and scenario selection. Then, after this was finished, the player would continue with the game as they normally would, typing each paragraph into the chat. Finally, the player would give the GPT their score and the GPT was instructed to act as the character who received the player's letter. The GPT was given additional instructions to be eccentric and to encourage interesting dialogue. Furthermore, it was instructed to not explain the rules of the game and not reveal the consequences of different scenarios to the player before the game had started.

Thousand Year Old Vampire

Thousand Year Old vampire is a game that played by creating a character from scratch; choosing a name, inventing a backstory, and assigning your vampire resources. Additionally, you must create 'mortals', who are NPCs who you will interact with during the game and memories which hold key events in the story. With each turn, the player must roll a d10 dice and a d6 dice, subtracting the d6 from the d10. In the game rules, there is a list of 100 prompts, and the result of the players' roll determines how many prompts they will move forward. Each prompt will give the player a decision to make. The player must make a decision, for example, sacrificing a mortal who was their friend, and then the player must narrate the story for themselves. This repeats until the player loses all resources or reaches prompt 100. There is complete freedom in this game, resources can be whatever the player wishes and the story can go in whichever direction they desire as long as it is in keeping with the prompt. For this prototype, the GPT was given the following instructions:

You will act as a narrator for a game of Thousand Year Vampire .

First ask the player to create a character for themselves, expand on the backstory if necessary.

During the game, when a prompt is received, you will not immediately start narrating. You will present the prompt to the player. They will answer the questions, and make the decisions in the prompt. Afterwards, you will narrate exactly how their decision plays out in great detail, make it exciting.

When a dice needs to be rolled, you will ask the player to do it. You will not make dice rolls yourself.

As can be seen, the GPT was instructed to act as the narrator of the story. It would supply the player with prompts and keep track of their resources, memories, mortals, etc. Furthermore, after receiving replies from the player, it was instructed to elaborate on their choices, creating an intricate story from them.

Four Against Darkness

Four Against Darkness is a dungeon crawler game in which the player creates a party of characters and explores a dungeon to kill monsters, find treasure, and level up. Players will construct the map of the dungeon as they move through it by rolling dice and consulting lookup tables for the size and shape of rooms they are adding. Additionally, dice rolls determine if loot, monsters, or other objects are in newly created rooms. Combat works by choosing attacks and rolling dice to determine damage. For this prototype the following prompt was given:

> *You will act as the game master for a game of Four Against Darkness. You will instruct the player how to play the game, helping them along each step. Remind them of their options. You will let players do the dice rolling rather than you doing it yourself. After a player makes a move you should narrate what is happening to them. Place a lot of emphasis on in depth narration when it is prudent. For example go into detail about the party as they enter a dungeon, describe the eeriness of the room, describe the monsters as they appear.*

As can be seen, this GPT is instructed to act as an assistant, helping players through the game as well as narrating as players move through the dungeon. The model is given particular emphasis on creating an atmosphere within the dungeon.

5.3 Prototype Evaluation

A focus group of the researcher and potential players was created to test the two prototypes. Within this group, each GPT was played multiple times, and adjustments were made to the prompts given to them to see how this might affect their outputs. Furthermore, small experiments were conducted on how the GPTs would respond to malicious input, such as asking them about AI. An open discussion was then held about the successes and failures of the GPTs, as well as which game would ultimately be used for the final product.

In terms of success, it was found that all the GPTs did well at keeping coherent and consistent thoughts throughout the trials. The narrative descriptions given by the GPT, particularly in the case of Thousand Year Old Vampire, were very detailed and exciting to read. Additionally, while playing Quill, the GPT had the possibility to produce very humorous outputs. Furthermore, with the inclusion of protective prompting, telling the GPTs not to respond to unrelated inputs, it was found to be quite difficult to affect the outputs of the models with malicious inputs.

Many considerations were also discussed. Firstly, a significant factor that was determined was the extent to which the user interface of the final product may influence immersion. This is significant because if a user interface is very complex, it may damage immersion; however, if intuitive, it could boost immersion. This would then undermine the results of the evaluation as it would not be clear whether the natural language interface or the user interface affected immersion. We, therefore, determined that the modified version's user interface should match the conventional

version's aesthetic as much as possible. To further decrease the uncertainty of what may be affecting immersion, it was also determined that a simple game may prove a stronger candidate. This is because if a player finds it simple to play a game, the user interface of the modified version may affect how the user plays less.

Another consideration that was discussed was the frequency of requests sent to the language model with respect to how long it takes to receive a response. It was pointed out that in a game like Quill, a request is not made very often, and it is therefore not immersion-breaking if it takes a while; however, in a game like Four Against Darkness, requests are made frequently, and thus may greatly influence immersion.

Eventually, a decision had to be made on which game would be used for the final product. Two key factors decided this. Firstly, while playing Thousand Year Old Vampire, it became evident that specific graphic themes were present in the game, and the GPT occasionally issued warnings when it outputted graphic content. As AI developers fall under the category ‘keep satisfied’ in the prior stakeholder analysis, it is of utmost importance that the regulations they set for their products must be respected. Therefore, Thousand Year Old Vampire was eliminated as a choice. Next, a discussion was held on the experience of Four Against Darkness. It was posited in this discussion that of all the games, this one was the most mechanically intensive and made it feel similar to a video game. All members of the group agreed upon this; however, whether or not this was a bad thing had to be considered. Ultimately, the decision was made to choose Quill as the final game, as in mechanical games, players may be less interested in narrative, and the complexity may affect evaluation results.

5.4 Final Specifications

After considering the various personas and evaluating the Lo-Fi prototypes, a final selection of specifications was created using MoSCoW. This can be found in Table 11.

Specification	Must	Should	Could	Won't
Target at least one form of immersion	X			
Target at least two forms of immersion		X		
Target all three forms of immersion			X	
Adaption of Quill	X			
Natural Language interface integration using a LLM	X			
NPC interaction	X			
Included rules	X			
Round time under 15 minutes	X			
Request response time of under 30 seconds	X			
Application aesthetics must imitate Quill aesthetics	X			
User-friendly interface		X		
Size under 2 GB		X		
Dice rolling feature		X		
Save feature			X	
Offline support			X	
Voice Input/Output			X	
Multi-Platform Support				X
Background Music				X

Table 11: Final Specifications

Chapter 6. Realisation

6.1 Tools

6.1.1 Environment

The Laptop SoloRPG program was created using Python 3.11.9 [40]. Python was chosen for this project for several reasons. Firstly, it is a highly versatile language that can be used for games, chatbots, and graphical user interfaces. Secondly, it has one of the largest user bases of all languages, meaning that support is more readily available if problems arise. Furthermore, Python has a massive collection of libraries that make many tasks easier. Finally, programming in Python takes a lot less time than other languages, making it ideal for creating quick prototypes until a final design is landed upon.

OpenAI Library

The OpenAI library [41] was used in the application to enable the program to make requests to GPT models through the OpenAI API. In order to use this API a connection between the program and a given GPT model is created. Next a prompt can be given to the model. This is done by creating a 'conversation history' in the form of an array of texts, with each text linked to a 'user profile'. For example, in the created application, the model is sent a conversation history with the first message being sent under the name 'System' informing the model of its task to respond to the player in a given way, and a second message under the name 'player' that contains the player's written text. After sending the request through the API, the program pauses and waits for a response, which is delivered through the API after a short delay.

TKinter Library

TKinter [42] is a GUI library used in this project to create a user interface for the players. Using the library, a window of a given size is created, and then buttons, images, and text can be added. Furthermore, additional windows can be created, and buttons can be linked to given methods which activated when the button is pressed.

6.1.3 Language Model

During the creation of the application, different language models were tested to find an optimal experience for players and those who would participate in the user evaluation. The first consideration was the time taken for a request to each model to return a response. It is important that upon finishing their letter, a player receives a response within an appropriate time frame. If, for example, a player had to wait for over a minute, they may become bored or even concerned that the application has broken, thus breaking immersion. The next considered factor was the size of the

program once a language model had been added. While some language models are hosted online and accessed through APIs, others are stored locally on device, and thus may substantially impact the size of the program. It is crucial that the total file size of the application is kept as small as possible because larger file sizes will cause longer download times, and users will require more storage space, which may be an issue for some users. Finally, the last consideration taken was the quality of responses, which varies depending on the complexity of a model. It is vital that the responses given by each model reach a certain threshold of coherence and relatedness to the player's input to guarantee immersion is not diminished.

Three models were therefore chosen to be tested: GPT-3.5 Turbo [43], GPT-4 [43], and Mixtral 8x7B [44]. Firstly, Mixtral 8x7B, a local model, was tested. Testing revealed that this model was able to give both coherent and quick responses to given player inputs, making it a good candidate. However, it was noted that the file size when including this model was over 8GB, and furthermore, response time may vary depending on the computational power of each user's computer. Next, GPT-4 was tested. This used the OpenAI API, meaning only the size of the OpenAI library was added to the project. Furthermore, this model was found to return coherent responses consistently and excelled at responding to the detailed contents of a player's letter. However, this model had by far the longest waiting time between sending a request and receiving a response. In some cases where a particularly long text was sent to the model, it could take over a minute to respond. Finally, GPT-3.5 turbo was tested. This model also used the OpenAI library and was much faster than both the Mixtral 8x7B and GPT-4 models. While the model was able to give coherent responses to player input, it did lack some of the detail and, in some cases, character of the other two models.

When comparing the three models against all factors, GPT-3.5 Turbo was judged to meet the combined considerations to the highest average degree and was therefore selected to be used in the final application.

6.2 Graphic User Interface

In order to create the graphic user interface, heuristics for designing graphical elements were first considered, and then a selection of programs was used to create the required assets.

During the LoFi prototype evaluation, a concern that differences in the visual experience between the original version and the modified version of the game may impact the player's immersion and thus may influence and ultimately reduce the validity of user evaluation. Therefore, where possible, images were taken directly from the original Quill game sheets and included in the modified version's corresponding elements. Furthermore, the general aesthetic of the original Quill

game sheets was replicated by attempting to match features such as fonts and colour palettes. A comparison between the profile of 'The King' in each version can be seen below in Figure 14.

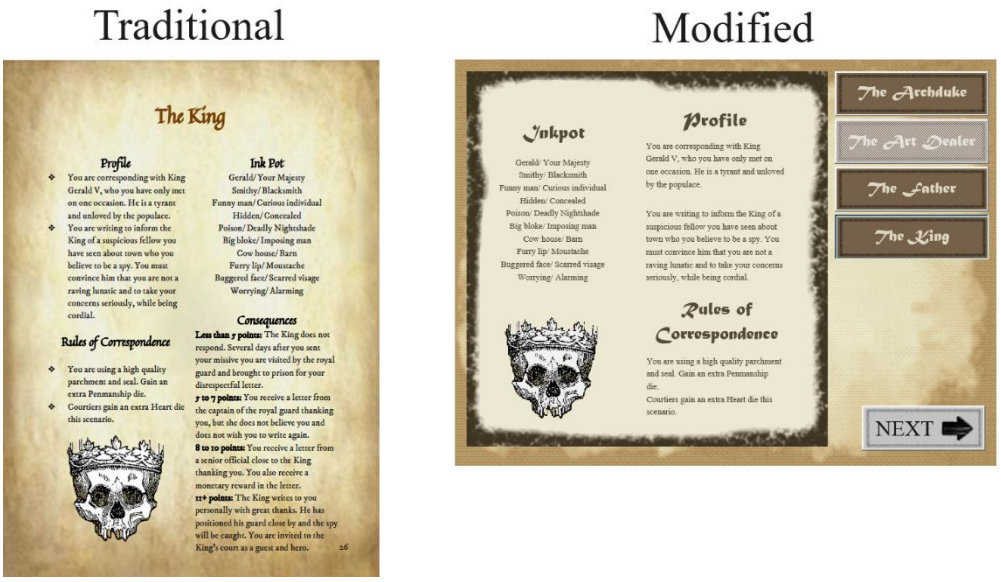

Figure 14: The King Scenario Comparison

In order to create the assets used in the program, two programs were used: Microsoft Paint [45] and Img2Pixel [46]. Microsoft Paint was used to create all images used in the game, including the background, profile images, button faces, etc. Each image was created in the resolution that would be used in-game, attempting to match the colour palette and fonts used on the game sheets as closely as possible. Once a base image was complete, it would then be modified using Img2Pixel. In this program, images can be adjusted in various ways, such as adding dithering to the image and confining it to a given colour palette, which helps match the aesthetic of the original version. A comparison between the base background made in Microsoft Paint and the background modified in Img2Pixel can be seen in Figure 14.

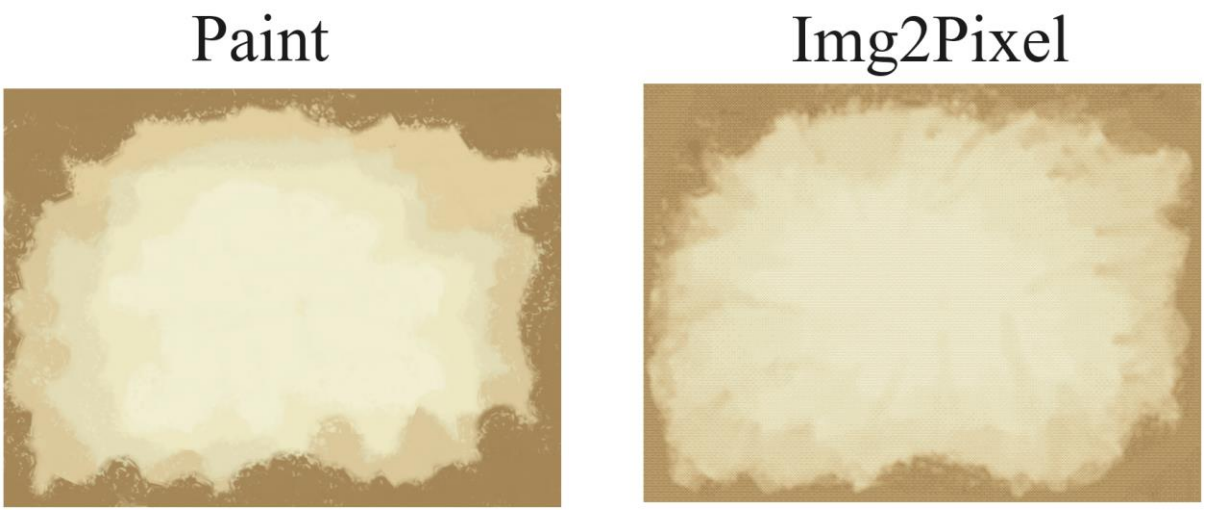

Figure 15: Paint versus Img2Pixel Comparison

6.3 Gameplay

To play, the player must first run a .exe file that begins the game program and opens a main window of 800x600 pixels in size. To allow for the player to play the game in a linear and easy-to-understand fashion, each stage of play was dissected into different 'screens'. These screens are displayed on the window as the game progresses and are as follows: profile and skill selection, scenario selection, letter-writing, and response reception. Additionally, a second window is created that displays the rules of the game so that players can understand their choices and how to play the game. The image displayed in the rules window can be seen in Figure 16.

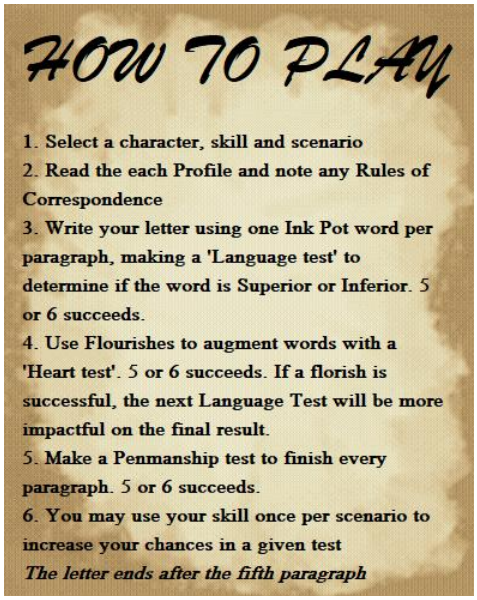

Figure 16: Rules Image

The modified version's gameplay followed the exact same flow as the traditional version. Players start by selecting a profile of the character they wish to roleplay and a skill, and after pressing

the ‘Next’ button, must choose the scenario they would like to correspond with. In each of these two selection screens, players must push buttons with the name of the profile or scenario they wish to view, upon which the corresponding profile/scenario image is displayed. An example of the player and skill select screen can be seen in Figure 17.

Figure 17: Character and Skill Select Screen

The next screen is the letter-writing screen, in which players compose a letter to their chosen correspondent. This page includes a description and inkpot of the selected scenario, five sections where each paragraph should be written, and various buttons for selecting words, using skills, and rolling tests. When a ‘heart’, ‘language’ or ‘flourish’ test button is pressed, a new window is created with the correct number of dice, which the player can roll using the ‘Roll’ button. Images of the text-writing and roll-dice windows can be seen below in Figure 18.

Figure 18: Letter-Writing Screen and Dice-Roll Window

Once the player has finished all 5 paragraphs, they can press the final 'Finish' button. After this, the program sends the player's letter to the language model, and upon receiving a response, the response reception screen is shown. On the screen, an image of a letter with the response given by the large language model is displayed for the player to read. They can then restart the game. An image displaying the screen and an example response can be seen in Figure 19.

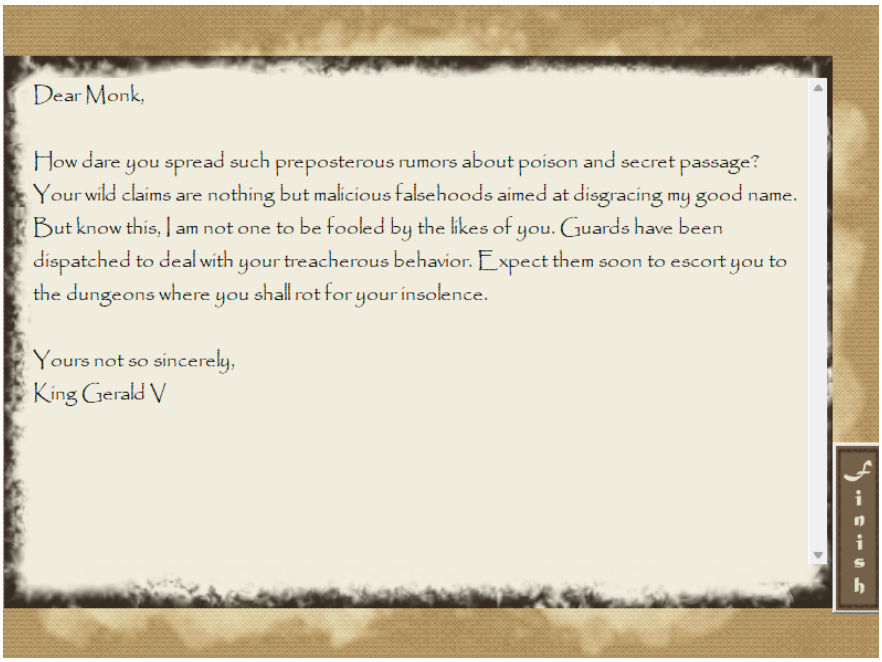

Figure 19: Example Response

Chapter 7. Evaluation

In this chapter the results of the conducted evaluation are examined and analysed to determine whether significance can be given to the findings.

7.1 User Evaluations

A total of thirty-six participants took part in the AB test conducted on the conventional and modified versions of Quill. Eighteen of the participants evaluated the traditional version first, and others evaluated the modified version first. Of the thirty-six participants, twenty-seven indicated that they had some experience with SoloRPGs prior. Throughout testing, there were very few problems with file sharing or users not understanding what they had to do.

Every user was able to complete the survey fully, playing both versions. After collecting all the responses, the results were firstly anonymised by removing data that could linked back to the participant, such as their name and the date of their evaluation. Next, the sums of the three categories, flow, presence, and absorption, were calculated for each participant and each version. Finally, the data was checked for significant outliers, however none were found.

7.2 Flow

In order to determine whether the inclusion of a large language model had an effect on immersion the total sum of flow per participant for the two versions was compared. This can be seen in Figure 20.

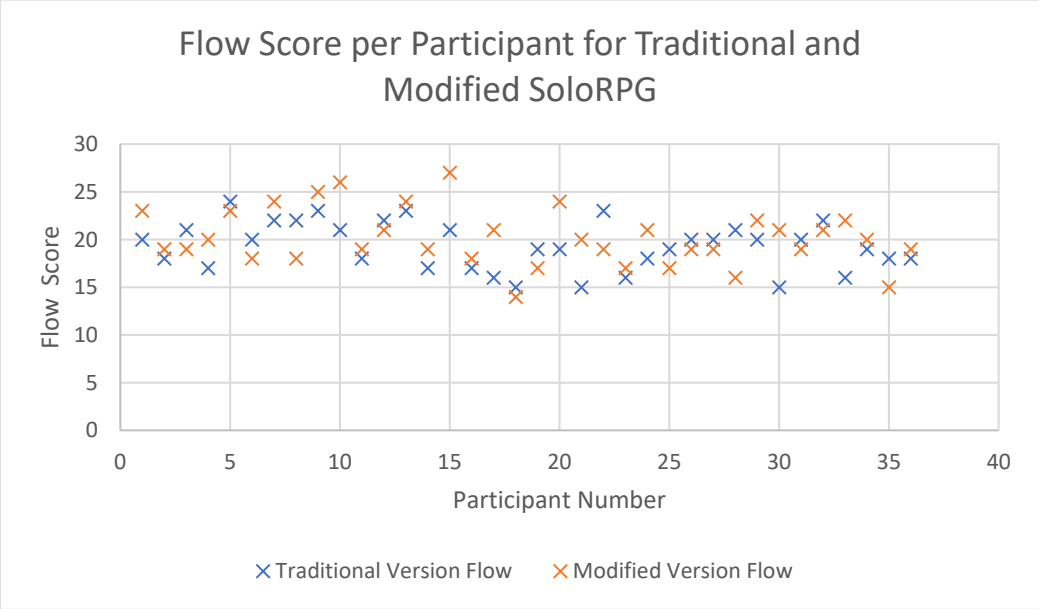

Figure 20: Flow Score per Participant for Traditional and Modified SoloRPG

In the graph above, it does seem that the modified version score is slightly higher on average compared to the scores for the traditional version, and this is confirmed as the mean for the modified version is 20.167, compared to 19.305 for the traditional version. Therefore, a two-tailed, paired T-test was performed to evaluate whether this is significant. The null and alternative hypotheses for this test are given below.

Null hypothesis (H_0): The mean difference between the paired observations is equal to zero

$$H_0 : \mu_d = 0$$

Null hypothesis (H_1): The mean difference between the paired observations is not equal to zero

$$H_1 : \mu_d \neq 0$$

Computing the t-test using Equation 1 yields a p-value of 0.09651. As the p-value is greater than the significance level of 0.05, the null hypothesis fails to be rejected. Thus, there is not sufficient evidence to conclude there is a significant difference in the flow experienced between the two versions.

7.3 Presence

To determine whether the inclusion of a large language model affected immersion, the total sum of presence per participant for the two versions was compared. This can be seen in Figure 21.

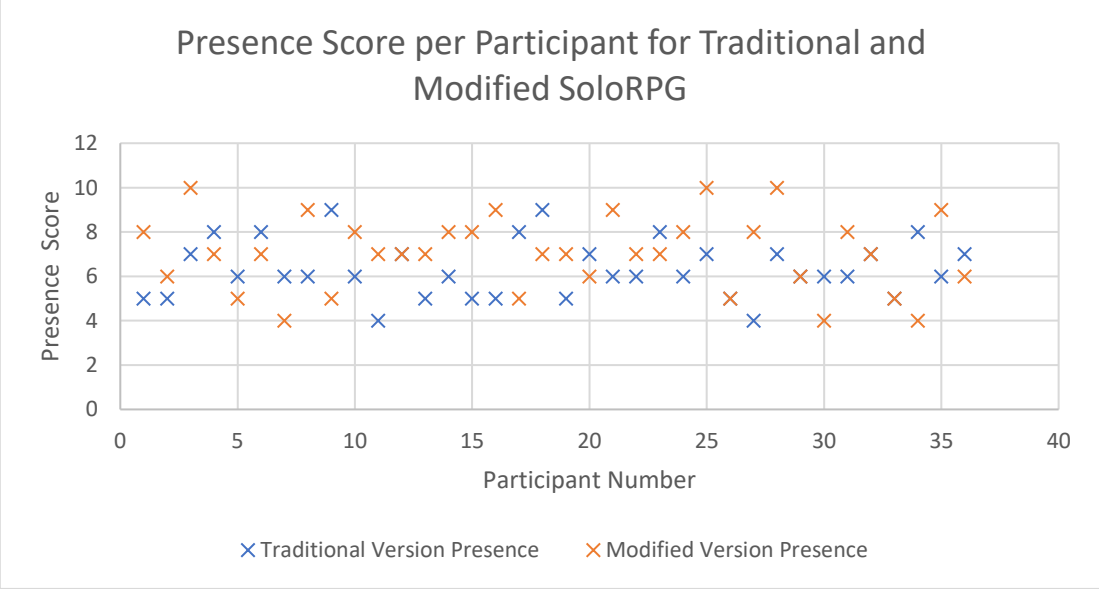

Figure 21: Presence Score per Participant for Traditional and Modified SoloRPG

In the graph above, the two versions seem to be equal. However, the mean for the modified version is 7.028, compared to 6.306 for the traditional version. Therefore, a two-tailed, paired T-test

was performed to evaluate whether this is significant. The null and alternative hypotheses for this test are given below.

Null hypothesis (H_0): The mean difference between the paired observations is equal to zero

$$H_0 : \mu_d = 0$$

Null hypothesis (H_1): The mean difference between the paired observations is not equal to zero

$$H_1 : \mu_d \neq 0$$

Computing the t-test yields a p-value of 0.0637. Again, as the p-value is greater than the significance level of 0.05, the null hypothesis fails to be rejected. Thus, there is not sufficient evidence to conclude there is a significant difference in the presence experienced between the two versions.

7.4 Cognitive Absorption

To determine whether the inclusion of a large language model affected immersion, the total sum of absorption per participant for the two versions was compared. This can be seen in Figure 22.

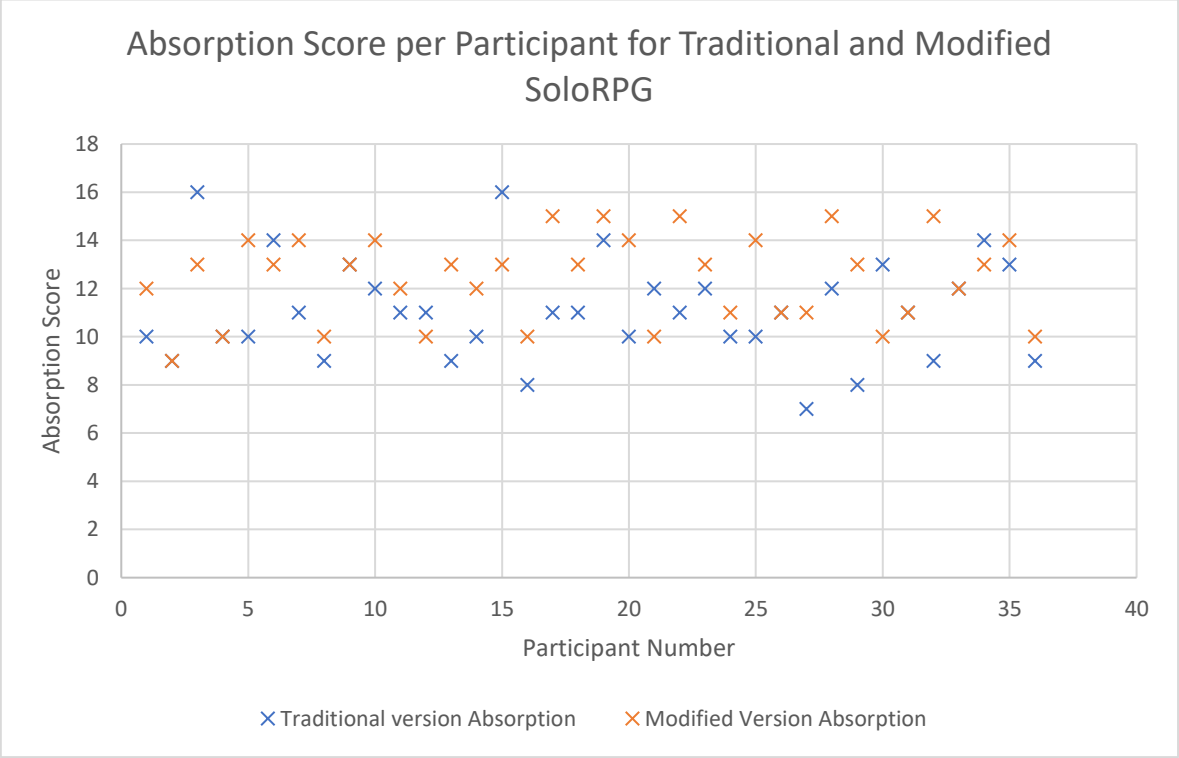

Figure 22: Absorption Score per Participant for Traditional and Modified SoloRPG

In the graph above, it does seem that the modified version score is quite significantly higher on average compared to the scores for the traditional version, and this is confirmed as the mean for the modified version is 12.417, compared to 11.083 for the traditional version. Therefore, a two-

tailed, paired T-test was performed to evaluate whether this is significant. The null and alternative hypotheses for this test are given below.

Null hypothesis (H_0): The mean difference between the paired observations is equal to zero

$$H_0 : \mu_d = 0$$

Null hypothesis (H_1): The mean difference between the paired observations is not equal to zero

$$H_1 : \mu_d \neq 0$$

Computing the t-test yields a p-value of 0.00147. As the p-value is less than the significance level of 0.05, the null hypothesis is rejected. Thus, there is sufficient evidence to conclude that there is a significant difference in the cognitive absorption experienced between the two versions.

7.5 Evaluation of Results

Although the results of the flow and presence score evaluations were not significant, the results of the absorption scores demonstrated that including a large language model increased player absorption. Therefore, as absorption is one of the three factors of immersion defined in this report, it can also be said that including a large language model in the Quill SoloRPG has enhanced player immersion per this paper's definition of immersion.

Chapter 8. Conclusion and Limitations

8.1 Conclusion

SoloRPGs provide a particular freedom not found in other games, bestowing narrative direction in a single player's hands, leading to intricate and complex scenarios that are entirely unique to each played round. Additionally, the experience of SoloRPGs largely relies on the immersion they can bring to a player. However, factors like complicated rulesets, many components, or messy tabletops can impede immersion. This presents the opportunity to employ a novel technology alongside a traditional SoloRPG to enhance player immersion while retaining narrative freedom. This project aimed to determine whether LLMs acting in a natural language interface had the potential to fill this role, as they demonstrate ample human-like conversational skills and can understand and respond to many complex situations. Therefore, the following research questions were identified:

RQ: "To what extent can player immersion in a SoloRPG be enhanced by a natural language interface using large language models?"

Sub RQ1: "What is player immersion?"

Sub RQ2: "What are the techniques used to measure player immersion?"

Sub RQ3: "How are large language models used to create natural language interfaces?"

To answer these questions, research was conducted, and a modified version of the SoloRPG Quill that implements LLM technology was created and tested against the traditional version. This modified version is played on the computer and mimics the original game as closely as possible. However, upon finishing a letter in the modified version of the game, the player is shown a response that OpenAI's GPT-3.5 Turbo model has generated. This differs from the traditional version, where the player must imagine the response.

A user evaluation was conducted that received 36 responses, in which players played both the original version of the game Quill and a modified version in a random order and completed a Likert scale questionnaire for each experience. Each question within this questionnaire fell into one of three categories: flow, absorption, or presence. In the categories of flow and presence, no significant difference was found between the two game versions. However, a significant difference was found in players' cognitive absorption, whereby players, on average, were significantly more absorbed in the modified version. This allows for the conclusion that player immersion was enhanced by implementing a natural language interface using a large language model. However, it is not possible to state anything beyond this. While this project aimed to create an experience that

contained the core elements of SoloRPGs in order to be able to make a general claim about enhanced immersion using natural language interfaces, there is not enough evidence. Therefore, it must be concluded that, to some extent, natural language interfaces using large language models can be used to enhance SoloRPG experiences. However, the extent to which this is true is unknown.

8.2 Limitations

The limitations of this approach must be considered to correctly evaluate this project and consider avenues for future research.

The first limitation of this project is its scope. To answer the research question, a modified version of the game Quill was created and tested against the original. As only one game is being used to compare modified and traditional SoloRPGs, it is difficult to make generalised claims based on evidence found during testing. It could be argued that the game Quill was carefully chosen to represent SoloRPG games, as the SoloRPG analysis that was performed found that Quill fit the core themes and contained the critical mechanics of an 'essential' SoloRPG experience. While this is true, there is so much variety within the genre of SoloRPGs that an essential experience does not include many different mechanics and elements that other, more outlying, SoloRPGs contain.

Another limitation of this evaluation is that the modified version of Quill may affect the player's experience in ways other than implementing a natural language interface. As the modified version is a SoloRPG that runs in a desktop application, specific measures had to be taken to ensure users would be able to play the game with as much ease as possible. For example, when the player begins the application, they are automatically taken to the character and skill select screen and, upon making a choice, are taken directly to the scenario-selection screen. This contrasts with the traditional version, where players may have to search through the game sheets in an unorganised manner to find all the information. Therefore, the user interface of the modified version may make playing the game more straightforward, which could, in theory, lead to enhanced player immersion and thus diminish the validity of the evaluation. However, an argument could be made that this user interface is part of or enabled by the natural language interface and, thus, does not affect the validity of the conclusion.

Other limitations may include factors related to the LLMs that were tested and used in this project. Factors such as latency, coherence, and creativity of the generated response may impact player immersion. Due to the limitations of current LLMs, player immersion may not be as significantly impacted as it would be with future models.

Chapter 9. Future Work

While the present study has made initial conclusions on enhancing player immersion in SoloRPGs through the integration of Large Language Models (LLMs), several avenues for future research remain. These future directions could help further understand and expand the potential applications of LLMs in gaming and interactive storytelling.

Firstly, a broader selection of games may improve findings. The current study focused on a single game, "Quill," to evaluate the impact of LLMs on player immersion. Future research could expand this to include a broader range of SoloRPGs, which may help determine if the observed enhancements in immersion are consistent across different game mechanics, narratives, and themes. By diversifying the selection of games, researchers can identify specific elements that benefit most from LLM integration.

Another avenue of future research could be to explore the impact of personalized AI responses on player immersion. By tailoring responses based on player behaviour, preferences, and past interactions, LLMs could create more engaging and individualized experiences. Investigating the effectiveness of different personalisation techniques could provide valuable insights into how players experience immersion and what factors affect this.

Additionally, as AI technology evolves, future studies should leverage AI understanding and creativity advancements. It will be crucial to investigate how more sophisticated models can create richer narratives, generate complex scenarios, and handle unexpected player inputs. This includes exploring the integration of multimodal AI systems that combine text, audio, and visual elements to create more immersive environments.

Finally, future research should explore novel ways to create natural language interfaces between humans and AI. In this project, a basic approach was taken, with players using a keyboard and mouse to communicate with the LLM. If, instead, this had been done with the player handwriting their letter and having a natural language interface that could read handwriting and send it to the model, results may differ.

By addressing these future research directions, the field can continue to enhance the immersive qualities of SoloRPGs and other interactive experiences, ensuring that AI technologies are utilised to their fullest potential while considering their broader implications.

References

[1] C. Jennet, A. L. Cox, P. Cairns, A. Epps, S. Dhoparee, T. Tijs and A. Walton, "Measuring and defining the experience of immersion in games," *International Journal of Human-Computer Studies,* vol. 66, no. 9, pp. 641-661, 2008.

[2] J. H. Brockymer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart and J. N. Pidruzny, "The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing," *Journal of Experimental Social Psychology,* vol. 45, no. 4, pp. 624-634, July 2009.

[3] H. Qin, P.-L. P. Rau and G. Salvendy, "Player Immersion in the Computer Game Narrative," *Entertainment Computing - ICEC,* pp. 458-461, 2007.

[4] M. Csikszentmihalyi, Flow: The Psychology of Optimal Experience, 1st, Ed., New York: Harper & Row, 1990.

[5] K. Procci and C. Bowers, "An Examination of Flow and Immersion in Games," in *PROCEEDINGS of the HUMAN FACTORS and ERGONOMICS SOCIETY 55th ANNUAL MEETING*, 2011.

[6] D. Weibel and B. Wissmath, "Immersion in Computer Games: The Role of Spatial Presence and Flow," *International Journal of Computer Games Technology,* vol. 2011, 2011.

[7] D. Yellowless and A. Hargadon, "The pleasure principle: immersion, engagement, flow," in *Proceedings of the eleventh ACM on Hypertext and hypermedia.*, New York, 2000.

[8] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Teleoperators and Virtual Environments,* vol. 7, no. 3, pp. 225-240, 1998.

[9] K. M. Lee, "Presence, Explicated," *Communication Theory,* vol. 14, pp. 27-50, February 2004.

[10] R. Tamborini and P. Skalski, "Playing Video Games," in *The Role of Presence in theExperience of Electronic Games*, New York, Routledge, 2006, pp. 225-240.

[11] S.-A. A. Jin, "I Feel Present. Therefore, I Experience Flow:" A Structural Equation Modeling Approach to Flow and Presence in Video Games," *Journal of Broadcasting & Electronic Media,* vol. 55, no. 1, 2011.

[12] R. Tamborini and N. D. Bowman, “Presence in Video Games,” in *Immersed in Media*, New York, Routledge, 2009.

[13] A. Tellegen, “Brief Manual for the Multidimensional Personality Questionnaire,” *Unpublished manuscript,* pp. 1010-1031.

[14] S. M. Roche and K. M. McConkey, “Absorption: Nature, assessment, and correlates,” *Journal of Personality and Social Psychology,* vol. 59, no. 1, pp. 91-101, 1990.

[15] R. Agarwal and E. Karahanna, “Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage,” *MIS Quaterly,* vol. 24, no. N, pp. 665-694, December 2000.

[16] A. Drachen, G. Yannakakis, L. Nacke and A. L. Pedersen, “Correlation between Heart Rate, Electrodermal Activity and Player Experience in First-Person Shooter Games,” *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games,* pp. 49-54, July 2010.

[17] D. J. Smith and N. T. Graham, “Use of eye movements for video game control,” in *Proceedings of the International Conference on Advances in Computer Entertainment Technology*, Hollywood, California, USA, June 2006.

[18] S. A. Jackson and H. W. Marsh, “Development and Validation of a scale to Measure Optimal Experience: The Flow State Scale,” *Journal of Sport and Exercise Psychology,* vol. 18, pp. 17-35, 1996.

[19] S. A. Jackson and R. C. Eklund, “Assessing Flow in Physical Activity: The flow state scale and dispositional flow scale,” *Journal of Sport and Exercise,* vol. 24, pp. 133-150, 2002.

[20] S. A. Jackson, A. J. Martin and R. Eklund, “Long and Short Measures of Flow: The Construct Validity of the FSS-2, DFS-2, and New Brief Counterparts,” *Journal of Sport and Exercise Psychology,* vol. 30, pp. 561-587, 2008.

[21] J. Hamari and J. Koivisto, “Measuring flow in gamification: Dispositional Flow Scale-2,” *Computers in Human Behavior,* vol. 40, pp. 133-143, 2014.

[22] B. G. Witmer and M. J. Singer, “Measuring Presence in Virtual Enviroments: A Presence Questionaire,” *Presence: Teleoperators and Virtual Environments,* vol. 7, pp. 225-240, 1998.

[23] A. Tellegen and G. Atkinson, “Tellegen Absorption Scale,” *APA PsycTests,* 1974.

[24] W. A. IJsselsteijn, A. W. de Kort and K. Poels, “The Game Experience Questionnaire,” *Technische Universiteit Eindhoven,* 2013.

[25] J. Dyer, “A Very Brief History of Gamebooks (up to 1979),” 6 June 2019. [Online]. Available: https://bluerenga.blog/2019/06/06/brief-history-of-gamebooks/. [Accessed 2024 Febuary 12].

[26] Chooseco, “CHOOSE YOUR OWN ADVENTURE,” [Online]. Available: https://www.cyoa.com. [Accessed 11 Febuary 2024].

[27] Fandom, “Fighting Fantasy,” [Online]. Available: https://fightingfantasy.fandom.com/wiki/Fighting_Fantasy. [Accessed 12 February 2024].

[28] J. Nuttall, “Original Scenarios Resurrected I: The Solo Dungeon (1977, Richard Bartle),” 21 June 2022. [Online]. Available: https://explorebeneathandbeyond.blogspot.com/2022/06/original-adventures-resurrected-i-solo.html. [Accessed 12 Februrary 2024].

[29] “AI Dungeon,” Latitude Inc, Oasis Tech Inc, 2019. [Online]. Available: https://aidungeon.com. [Accessed 02 05 2024].

[30] OxfordLanguages, “Oxford Languages and Google,” [Online]. Available: https://languages.oup.com/google-dictionary-en/. [Accessed 1 February 2024].

[31] Rabbit, “rabbit os,” rabbit inc, 2024. [Online]. Available: https://www.rabbit.tech/rabbit-os. [Accessed 02 05 2024].

[32] rabbit, “Introducing r1,” 9 January 2024. [Online]. Available: https://www.youtube.com/watch?v=22wlLy7hKP4. [Accessed 02 May 2024].

[33] “Ai Pin - Wearable Ai,” Humane Inc, 2024. [Online]. Available: https://humane.com. [Accessed 02 05 2024].

[34] E. Smith, “Interface Interference,” Tedium, 2024 April 30. [Online]. Available: https://tedium.co/2024/04/30/rabbit-r1-humane-ai-hardware-design-criticism/. [Accessed 10 May 2024].

[35] Coffeezilla, “Rabbit Gaslit Me, So I Dug Deeper,” YouTube, 24 May 2024. [Online]. Available: https://www.youtube.com/watch?v=zLvFc_24vSM. [Accessed 24 May 2024].

[36] A. Griffin, “Rabbit R1: Reviews of widely-hyped AI device say it is a ‘mess’ and not ready,” Independant, 30 April 2024. [Online]. Available: https://www.independent.co.uk/tech/rabbit-r1-review-round-up-ai-artificial-intelligence-b2537491.html. [Accessed 10 May 2024].

[37] A. Mader and W. Eggink, “A Design Process for Creative Technology,” *International Conference on Engineering and Product Design Education,* vol. 16, 2014.

[38] A. L. Mendelow, “Environmental Scanning -- The Impact of the Stakeholder concept,” *ICIS 1981 Proceedings,* vol. 20, 1981.

[39] D. Clegg and R. Barker, Case Method Fast-Track: A Rad Approach, Boston: Addison-Wesley Longman, 1994.

[40] Python, “Python 3.11.9,” 2 April 2024. [Online]. Available: https://www.python.org/downloads/release/python-3119/. [Accessed 10 June 2024].

[41] OpenAI, “Libraries,” [Online]. Available: https://platform.openai.com/docs/libraries. [Accessed 10 June 2024].

[42] Python Software Foundation, “tkinter — Python interface to Tcl/Tk,” 2 June 2024. [Online]. Available: https://docs.python.org/3/library/tkinter.html. [Accessed 20 June 2024].

[43] OpenAI, “Models,” [Online]. Available: https://platform.openai.com/docs/models. [Accessed 10 June 2024].

[44] Mixtral AI, “Open source models,” [Online]. Available: https://mistral.ai/technology/#models. [Accessed 5 June 2024].

[45] Microsoft, “Paint makes it easier to create,” [Online]. Available: https://www.microsoft.com/en-us/windows/paint. [Accessed 1 June 2024].

[46] L. Holzbeierlein, “SLK_img2pixel,” 20 January 2021. [Online]. Available: https://captain4lk.itch.io/slk-img2pixel. [Accessed 1 June 2024].

[47] R. M. Banos, C. Botella, M. L. A. Raya and V. Liano, “Immersion and Emotion: Their Impact on the Sense of Presence,” *Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society,* pp. 734-741, January 2005.

[48] W. Wirth, T. Hartmann, S. Bocking and P. Vorderer, “A Process Model of the Formation of Spatial Presence Experiences,” *Media Psychology,* vol. 9, no. 3, pp. 493-525, May 2007.

[49] M. Slater, M. Usoh and A. Steed, “Depth of Presence in Virtual Environments,” *Presence: Teleoperators and Virtual Environments,* vol. 3, no. 2, pp. 130-144, 1994.

[50] P. Zahorik and R. L. Jenison, “Presence as Being-in-the-World,” *Presence,* vol. 7, no. 1, pp. 78-89, February 1998.

[51] Y. Yao, J. Duan, K. Xu, Y. Cai, Z. Sun and Y. Zhang, “A Survey on Large Language Model (LLM) Security and Privacy: The Good, The Bad, and The Ugly,” *High-Confidence Computing,* vol. 4, no. 2, 2024.

Appendices

Appendix A: Game Engagement Questionnaire

1. I lose track of time (Absorption)
2. Things seem to happen automatically (Presence)
3. I feel different (Absorption)
4. I feel scared (Absorption)
5. The game feels real (Flow)
6. If someone talks to me, I don't hear them (Flow)
7. I get wound up (Flow)
8. Time seems to kind of stand still or stop (Absorption)
9. I feel spaced out (Absorption)
10. I don't answer when someone talks to me (Flow)
11. I cannot tell that I'm getting tired (Flow)
12. Playing seems automatic (Flow)
13. My thoughts go fast (Presence)
14. I lose track of where I am (Absorption)
15. I play without thinking about how to play (Flow)
16. Playing makes me feel calm (Flow)
17. I play longer than I meant to (Presence)
18. I really get into the game (Overall Immersion)
19. I feel like I just can't stop playing (Flow)