DM

Leveraging LLM Generated Narratives in Visual Novel Games for Mental Health Literacy

MASTER DISSERTATION

José Pedro Vieira Sousa
MASTER IN INFORMATICS ENGINEERING

UNIVERSIDADE da MADEIRA
A Nossa Universidade
www.uma.pt

September | 2025

Leveraging LLM Generated Narratives in Visual Novel Games for Mental Health Literacy

MASTER DISSERTATION

José Pedro Vieira Sousa
MASTER IN INFORMATICS ENGINEERING

SUPERVISOR
Pedro Filipe Pereira Campos

CO-SUPERVISOR
Paulo Alexandre Câmara Bala

Faculty of Exact Sciences and Engineering

Master of Science Degree in Informatics Engineering

Leveraging LLM Generated Narratives in Visual Novel Games for Mental Health Literacy

José Pedro Vieira Sousa

Supervised by:

Prof. Dr. Pedro Filipe Pereira Campos

Prof. Dr. Paulo Alexandre Câmara Bala

Constituição do júri de provas públicas:

Prof. Dr. Monica da Silva Cameirão, Presidente

Prof. Dr. Fabio Ruben Silva Mendonça, Vogal

Prof. Dr. Paulo Alexandre Câmara Bala, Vogal

Funchal, November 2025

Resumo

A saúde mental, apesar de ser um componente crucial da saúde, tem sido frequentemente negligenciada e estigmatizada. Os estudantes universitários, devido às suas situações e aos seus factores de stress únicos, são mais vulneráveis a desafios de saúde mental e ao estigma. A crescente incidência de problemas de saúde mental, agravada em parte pela pandemia da COVID-19, expôs as consequências significativas e falhas dos sistemas de apoio à saúde mental negligenciados. Para atenuar estas falhas, houve um aumento na investigação das intervenções digitais de saúde, como uma alternativa às abordagens tradicionais, não só na educação sobre a saúde mental, mas também no diagnóstico, monitorização e cuidado dos indivíduos. As intervenções digitais de saúde mental, como os jogos, podem ser escaláveis, sob demanda e acessíveis. Este trabalho alinha jogos (como romances visuais) com modelos de linguagem de grande escala para criar uma intervenção digital com o objectivo de aumentar a literacia em saúde mental. Ao combinar características dos jogos como interatividade, narrativa e estética, a intervenção digital proposta pretende permitir ao jogador explorar uma narrativa sobre a depressão, aprendendo sobre o papel das redes de apoio e cuidado. Realizamos um estudo piloto com cinco profissionais de saúde mental usando um prompt-game gerado por um modelo de linguagem de grande escala, para avaliar as narrativas geradas e informar o design do prototipo final. Através de entrevistas semiestruturadas sobre o protótipo e a narrativa, identificamos vários problemas, como a falta de informações ou a simplificação excessiva de conceitos. Apesar das limitações, os participantes consideraram o protótipo utilizável e potencialmente valioso para uso profissional. Com base nos resultados do estudo piloto, desenvolvemos um jogo de novela visual, usando modelos de linguagem de grande escala para a geração de narrativas, testado através de uma pesquisa de métodos mistos com 28 estudantes universitários, avaliando o impacto do jogo na literacia e nos falsos conceitos sobre depressão (avaliados antes, depois do teste e 1 semana depois), e o grau de transporte narrativo da narrativa gerada. Os resultados não apresentaram mudanças significativas na literacia da depressão, possivelmente devido aos altos níveis de conhecimento base. No entanto, os falsos conceitos sobre depressão diminuíram significativamente e mantiveram-se no seguimento. Os participantes relataram níveis moderados a altos de transporte narrativo, sugerindo forte envolvimento com a narrativa. Os resultados qualitativos realçaram o envolvimento emocional, a redução do estigma, o valor percebido das ferramentas

digitais, e o ceticismo dos participantes em relação à autenticidade das narrativas geradas por Inteligência Artificial. Estes resultados sugerem que narrativas interativas geradas por Inteligência Artificial podem ajudar a reduzir o estigma e os falsos conceitos sobre depressão, sendo uma possível ferramenta para intervenções educacionais.

Keywords: Saúde Mental · Transtornos Depressivos · Modelos de Linguagem de grande escala · Narrativas Interativas · Jogo

Abstract

Mental health, while a crucial component of health, has often been neglected and stigmatised. College students, due to their unique situations and stressors, are more vulnerable to experiencing mental health challenges and stigma. The increasing incidence of mental health conditions, exacerbated in part by the COVID-19 Pandemic, has exposed significant consequences and flaws of neglected mental health support systems. To counteract these flaws, digital health interventions have seen increased research as an alternative to traditional approaches, not only in educating individuals about mental health, but also in diagnosing, monitoring, and caring for them. Digital mental health interventions, such as games, can be scalable, on-demand, and affordable. This work aligns games (such as visual novels) with large language models to create a digital intervention aimed at increasing mental health literacy. By combining characteristics of games such as interactivity, narrative and aesthetics, the proposed intervention aims to give players agency to explore a narrative about depression, learning about the role of support networks and care. We conducted a pilot study with five mental health professionals using a large language model-generated prompt game to assess the generated narratives and inform the design of the final prototype. Through semi-structured interviews about the prototype and the narrative, we identified several problems, such as missing information or oversimplified concepts. Despite its limitations, participants viewed the prototype as usable and potentially valuable for professional use. Informed by the findings from the pilot study, we developed a visual novel game using large language models for the narrative generation, tested through a mixed methods research with 28 college students, evaluating the impact of the game in the depression literacy and misconceptions (evaluated before, after the play test, and 1 week later), and the degree of narrative transportation of the generated narrative. Results showed no significant change in depression literacy, likely due to high baseline knowledge. However, misconceptions of depression significantly decreased and were maintained at follow-up. Participants reported moderate to high levels of narrative transportation, suggesting strong engagement with the story. Qualitative findings emphasised emotional engagement, stigma reduction, perceived value of digital tools, and scepticism regarding Artificial Intelligence-generated narratives authenticity. These findings suggest that interactive, Artificial Intelligence-generated narratives can help reduce stigma and misconceptions of depression, offering a possible tool for educational interventions.

Keywords: Mental Health · Depressive Disorder · Large Language Models · Interactive Storytelling · Game

Acknowledgements

First, I would like to thank my supervisors, Professor Pedro Campos and Professor Paulo Bala, for their guidance, support, feedback, and the many lessons learned throughout this project.

I also wish to thank my friends Margarida, Carolina and Vera for their encouragement, support, and the many conversations and moments we shared in coffee shops that helped me along this journey, as well as my colleagues from eGamesLab for their support and the conversations we shared.

To Dr. Carlos, for the constant help over the years, which allowed me to better understand myself and to develop greater literacy about depression and mental health.

To Teacher Idalina, for the support and guidance during one of the moments of life when it was most needed.

I would also like to thank eGamesLab for funding through the Portuguese Recovery and Resilience Program (PRR), IAPMEI/ANI/FCT under Agenda no.26, C645022399-00000057.

Lastly, I could not fail to thank my family – my mother, father, brother, grandparents, uncles, aunts, and godparents – for their support, patience, and encouragement throughout this entire journey, and throughout my life – to everyone who made it possible for me to believe and to arrive here today. My sincerest thanks.

Table of Contents

List of Figures

List of Tables

List of Listings

Acronyms

AI Artificial Intelligence. 2, 12, 15, 38, 39, 60, 63, 66

APA American Psychological Association. 6, 7, 9

API Application Programming Interface. 32, 38–40, 46, 47

App Application. xi, 11, 29, 33, 38, 51

BLoC Business Logic Component. 15, 35, 43, 44, 46, 47

CA Conversational Agent. 4, 12, 13

CDC Centers for Disease Control and Prevention. 6

CS College Student. 1–4, 6–8, 16, 18, 49, 59, 63

DepDis Depressive Disorder. 1, 6, 7

DepSter Scale Depression Literacy and Misconceptions Scale. 51, 52, 54, 66

DL Depression Literacy. 3, 49, 51–56, 58, 63, 64, 66

DSM-5 Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. 15, 17, 19, 23

HCI Human Computer Interaction. 4

LLM Large Language Model. 2–4, 13–17, 19–21, 23, 29, 38–41, 43, 46, 48, 51, 64, 66

MH Mental Health. 1–13, 15–19, 21–23, 49, 59, 63, 64, 66

MHL Mental Health Literacy. 1, 3, 4, 7, 8, 10–12, 15, 17–19, 49, 51, 58, 59, 64, 66

MiscD Misconceptions about Depression. 3, 49, 51–58, 63, 64

NT Narrative Transportation. 51, 52, 56, 58, 63, 66

NT Scale Narrative Transportation Scale. 51, 52, 54

PE Prompt Engineering. 2, 4, 13, 14, 19, 24

SN Support Network. 2, 4, 6, 7, 16–18

TDD Test Driven Development. 15, 35, 43, 44

UI User Interface. 31, 32, 35, 38, 39, 43–47

VCA Voice-based Conversational Agent. 13

VN Visual Novel. 2–4, 15, 38, 43, 49, 59, 63

WHO World Health Organization. 1, 5, 6

1 Introduction

1.1 Problem and Motivation

With the creation of the World Health Organization (WHO) in 1948, the definition of health extended to recognise Mental Health (MH) as an essential aspect of overall well-being [1]. MH is defined as the ability to handle stress, develop one's abilities, and contribute to society, becoming a growing focus of research across disciplines [1]. MH systems suffer from several problems, some made evident or further amplified in recent years with the COVID-19 pandemic; specifically, a surge in service demand while the lack or limited accessibility of services remain, due to an increase in incidence and severity of MH conditions [2]. While existing problems related to MH persist, such as stigma and misinformation, they continue to limit accessibility to services [3, 4]. After COVID-19, there was a total increase of 28% in Depressive Disorders (DepDis) in one year [2, 5]. DepDis are more prevalent in adults (5% prevalence for people aged 20 years or older) and rare in people under the age of 10, having seen an increase in incidence and severity, especially among young adults (who experience higher stress levels than older generations and a higher prevalence change compared to these older groups), resulting from the impact of school and social restrictions [2, 6]. While these problems affect all age groups, College Students (CSs) go through a stressful period of transition to adulthood and the start of college. In addition, they also face unique stressors, including academic pressure, financial concerns, family issues, personal psychological issues, and college life adjustment challenges [6–10], turning into a period during which MH conditions commonly emerge, with 75% of cases beginning before age 24 [5, 11, 12].

Mental Health Literacy (MHL) has an important role, in helping CSs in controlling and preventing MH problems for themselves, and in helping others [11]. In addition, previous work by Lam [13] shows that lower MHL increased the risk of having higher severity level of depression, and with a probability of approximately 60% of suffering depression as opposed to the ones that have adequate MHL. Younger individuals often struggle to identify symptoms, as recognition improves with age, while misconceptions (such as interpreting symptoms as temporary age-related crisis) and fears about stigma and discrimination, and concerns about confidentiality further discourage seeking help [11,14,15]. Intersecting MH with technology has the potential to address these issues by creating interventions for treatment, awareness, or prevention of MH conditions [16]. More

specifically, and related to the lack or limited accessibility of services, there has been an increase in the use of digital MH interventions or tele-mental health support [2]. While existing work has focused on using technology on supporting existing MH practitioners [16, 17], recent studies explore alternative interventions using games [16–18] or Artificial Intelligence (AI) [19–21]. Although not yet clinically validated [16], these exploratory artefacts have expanded current intervention options; as such, this work engages with the design space of games, AI, and MH. Educational tools remain under-represented even as evidence suggests that embedding psycho-educational content within interactive formats enhances engagement and retention [16–18, 22, 23]. One type of game where the narrative design is particularly important is Visual Novels (VNs), as this type of game is more narrative-driven. Camingue et al. [24], bridging academic definitions and existing games, offer this definition of VNs as:

> *"[...] a digital narrative focused game that requires interactions where the player must be able to impact the story world or the story's progression. The story and interactions are most commonly presented through a text box and often employ additional forms of interaction including menu choices — which often contain sets of actions that the player character can perform — or dialogue options representing the player character's speech or thoughts. [...] The aesthetics of VNs are most often conveyed through static images of characters, background art, sound effects (SFX) feedback, and soundtracks."* [24]

Thus, VNs, defined as interactive stories driven by the player choices [24], using advances in Large Language Models (LLMs) and Prompt Engineering (PE) [25, 26], can be used to generate context-sensitive, dynamically generated personalized narratives reflecting the players' experiences and promote mental health [19–21, 27, 28]. These tendencies show the need for targeted interventions during this period. Conversely, strong social Support Networks (SNs) have been shown to improve MH outcomes, with a meta-analysis reporting a moderate positive correlation (effect size 0.337) between social support and student well-being [29].

1.2 Objectives and Research Questions

Following a research through design approach [30], and spurred by recent efforts in the incorporation of LLMs in systems [25], this work explored the incorporation of LLMs into VN games. Before creating the prototype and evaluating it with CSs, we considered the ethical ramifications of

designing digital MH interventions and performed a pilot study with five MH professionals. This study used a low-fidelity prototype, specifically a prompt-based narrative game, run on ChatGPT, where it generates the game narrative through the use of rules for the prompt game and guidelines for the narrative, focusing on the CSs' life events and stressors that can affect their MH. While we were not looking for clinical validity, expert feedback highlighted design considerations that informed the development of the main prototype (Section 4.1.1) [31]. The pilot confirmed LLMs' potential for generating content in sensitive contexts and suggested that it may be possible to expand this concept to other MH topics and topics beyond the MH realm.

Based on findings from the pilot study, the prototype of the VN game uses an LLM to dynamically generate story segments through the use of rules and guidelines, with the theme focusing on the CSs' life events and stressors that can affect their MH, specifically depression-related MH. This prototype engages players in a narrative where they are encouraged to choose actions as if they were the character in the story, thus seeing possible outcomes and aimed at improving their Depression Literacy (DL) while also reducing Misconceptions about Depression (MiscD). A user study with 28 CSs was conducted to evaluate the prototype's effectiveness in enhancing MHL and its narrative engagement. Through the use of this mixed-methods research, we apply the use of scales and a semi-structured interview to address the following research questions:

- **RQ1** – *How can we leverage large language models to generate narratives in game-based mental health interventions?*
- **RQ2** – *How do game-based mental health interventions affect the mental health literacy of young adults?*

1.3 Findings and Contributions

The findings obtained from the mixed methods study showed a significant decline in MiscD among participants, suggesting the intervention had the capacity to correct false beliefs. Nevertheless, no significant effects were observed in DL, possibly due to the high baseline knowledge levels among participants. The qualitative findings stressed that LLM-generated narratives were perceived as sensitive, relatable and realistic, contributing to a sense of reflection and immersion. With participants also expressing interest in the potential to expand this approach for other MH topics beyond depression.

Its contributions to Human Computer Interaction (HCI) are twofold:

- Firstly, this research contributes with the design and development of an artefact, specifically a VN game about depression, driven by an LLM, and controlled through prompt guidelines designed to increase MHL, depicting realistic student experiences with depression;
- Secondly, this research contributes with, user evaluations of the gamified health intervention with participants from formal and informal SNs, this evaluation will focus on the feasibility, impact, and limitations of using LLM-based interactive storytelling for MH education, particularly in reducing misconceptions and enhancing engagement.

Adding to these contributions, this work also resulted in two academic articles. The first, an academic poster for the pilot study with MH professionals, was presented at the International Academic Mindtrek Conference[1] and published in October of 2024 [31]. The second, a full article of the experiment with CSs, was accepted in the 2025 18th International Conference on Interactive Digital Storytelling (ICIDS 2025)[2], to be published in January 2026[3].

1.4 Document Structure

The remainder of this document is divided in 9 chapters. Chapter 2, "Theoretical Background", will discuss the psychological theoretical background needed in this work, including information about "Mental Health", "Depressive Disorders", "Support Networks" and "Stigma and Awareness of Mental Health Conditions". Chapter 3 presents the Computer Science background related to this work, informing about "Games as Digital Health Interventions", "Conversational Agents and Prompt Engineering", "Large Language Models as Game Agents" Following the previous chapters, Chapter 4 presents the design of the prototypes, including the pilot prompt game and the design of the prototype of this project. Chapter 5 presents the implementation of the prototype following the design considerations from Chapter 4. Chapter 6 presents the study design, protocol, measures used and the analysis of the data. Chapter 7 reports the results obtained for both quantitative and qualitative methods. The "Discussion" is given in Chapter 8, and the "Conclusion" is found in Chapter 9.

[1] https://www.mindtrek.org/academic-mindtrek/
[2] http://icids2025.ardin.online
[3] https://doi.org/10.1007/978-3-032-12405-0_12

2 Theoretical Background

This Chapter highlights key concepts and terms around the topic of MH, helping to contextualise the application domain.

2.1 Mental Health

MH has been recognised as an essential aspect of health. The WHO's Constitution defines health as:

> "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." [1]

This definition acknowledges MH as an essential component of health since WHO's creation in 1948. With WHO defining MH as:

> "A state of mental well-being that enables people to cope with the stresses of life, to realize their abilities, to learn well and work well, and to contribute to their communities. Mental health is an integral component of health and well-being and is more than the absence of mental disorder." [2]

WHO highlights in its "World Mental Health Report" that MH is not defined as a distinct value (e.g., healthy or ill, existence or absence of a MH condition) but rather as a value in a range, where a person may have a MH condition and be mentally healthy or mentally ill, the existence of one not implying the other [2].

In this work, we use the term *"mental health condition"* to refer to a subset of conditions, encompassing mental disorders, mental states, and psychological disabilities, the term is defined by WHO as:

> "A broad term covering mental disorders and psychological disabilities. It also covers other mental states associated with significant distress, impairment in functioning, or risk of self-harm." [2]

MH is influenced by several types of factors or determinants [32], including:

- biological factors – like genetics and physical health, that are connected to the biology of the individual;

- psychological factors – like emotional states and personality traits, that concern the psychology of the individual;
- social factors – like individuals that you contact directly or indirectly, or SNs, that are connected to the environment around the individual.

These factors affect the development of MH conditions and also the reactions on how people cope with them and seek help.

MH concerns are prevalent among high school and CSs and have been researched in recent years, with the COVID-19 pandemic taking a toll on the MH system, with students being some of the affected groups. Particularly high school and CSs, with the study "Stress in America 2020: A National Mental Health Crisis" [33] reporting that adults aged 18 to 23 years old, had the highest average stress levels (6.1 in 10), compared to adults [33]. The survey "Youth Risk Behaviour Survey" from the Centers for Disease Control and Prevention (CDC) [34], has shown the MH status of high school students, reporting that: 40% "Experienced Persistent Feelings of Sadness or Hopelessness"; 29% "Experienced Poor Mental Health During the Past 30 Days"; 20% "Seriously Considered Attempting Suicide During the Past Year"; 16% "Made a Suicide Plan During the Past Year"; 9% "Attempted Suicide During the Past Year"; 2% were "Injured in a Suicide Attempt During the Past Year". For CSs the 2022-2023 "The Healthy Minds Program" study [35] shows that: 20% had "Severe Depression"; 20% had "Moderate Depression"; 41% had "Any Depression"; 17% had "Severe Anxiety"; 19% had "Moderate Anxiety"; 36% had "Any Anxiety". 14% had "Suicidal ideation (past year)"; 6% had a "Suicide plan (past year)"; 2% made a "Suicide attempt (past year)"; 29% had a "Non-suicidal self-injury (past year)".

These values show the consequences of MH, and the need to address MH in educational environments, also showing the importance of increasing: early identification, awareness and support, for students' MH.

2.2 Depressive Disorders

DepDis are one of the most prevalent mental disorders, with 28.9% of people globally living with DepDis in 2019, from a total of 970 million that have mental disorders [2]. DepDis are defined by WHO as a *"common mental disorder"*, that *"involves a depressed mood or loss of pleasure or interest in activities for long periods of time"* [36]. And by American Psychological Association (APA) as *"any of*

the mood disorders that typically have sadness or empty or irritable mood as the predominant symptom" [37]. Highlighting that one of the most important symptoms of DepDis are mood changes, which occur with some level of permanence and in all areas of life, compared to normal mood changes that are temporary and related to certain areas of life that are connected to the situation that originated the mood change [38].

DepDis are more prevalent in adults, with a 5% prevalence for people aged over 20 years old, while being rare in people under 10, with a 0.1% prevalence between the ages of 5 and 9. No data is reported for children under 5 years [2]. A study from "Global Burden of Disease" showed that after COVID-19, there was an increase of 28% in DepDis in 2020 [2]; additionally, this increase affected younger people more compared with older groups, a possible result of the impact from school and social restrictions [2].

Considering the prevalence and impact of DepDis, especially on young adults, these are a critical focus of this work. Understanding the characteristics, symptoms and causes is needed to design interactive interventions aimed at increasing MHL and reducing stigma related to depression, in CSs, considering their unique situations and stressors.

2.3 Support Networks

SNs (e.g., friends, family) are one of the social factors that can greatly influence and contribute to overall MH, as well as help decrease stigma about MH. A good SN is important as they not only directly affect the MH of the individual, but can also intervene to offer care and/or guidance to other support resources [39]. In the APA Dictionary of Psychology, APA defines this social support given by people of an individual's social network as *"the provision of assistance or comfort to others, typically to help them cope with biological, psychological, and social stressors"* [40]. It also indicates that support can be provided by any relationship in the individual's social network in various forms, including practical help, tangible support, or emotional support [40]. The survey "Stress in America: The Impact of Discrimination" by APA in 2015 [6] shows that individuals with emotional support have an average of 5.0 points on a 10-point scale regarding stress levels, lower than individuals without emotional support who have an average of 6.3 points. However, the existence of social support on its own is not the only factor that influences the MH of the individual, as the individual needs to believe that they receive social support in any of its forms [29]; if this

belief isn't present, then the social support may not even be perceived by the individual. This is even more important for determined age groups, "Stress in America: The Impact of Discrimination" found that young adults have in average higher stress levels than older generations [6], in fact a meta-analysis from 2017 [29] concerning the correlation between social support and MH supports this by showing that social support on the MH of CSs had an effect size *"higher than the average level"* (0.337), with CSs having a higher access rate to social networks and as a consequence increasing the chances of looking for social support and skills in receiving it.

2.4 Stigma and Awareness of Mental Health Conditions

MHL defined by Jorm et al.'s [41, 42] as "*knowledge and beliefs about mental disorders which aid their recognition, management or prevention*", identifies six components of MHL, defined as follows: the "*ability to recognize specific disorders or different types of psychological distress*"; the "*knowledge and beliefs about risk factors and causes*"; the "*knowledge and beliefs about self-help interventions*"; the "*knowledge and beliefs about professional help available*"; the "*attitudes which facilitate recognition and appropriate help-seeking*"; and the "*knowledge of how to seek mental health information*". Through this definition, we can see that for someone to have a certain degree of MHL, there's a need to have a conjunction of several components, e.g., not only knowing about the symptoms but also knowing the risk factors and causes, how to recognise them, and how to seek help. Thus, defining MHL as a conjunction of the knowledge that someone needs, so that they can help themselves and others, either by recognising, managing or preventing. However, Jorm [43] later noted that there have been several works on MHL, though most focus only on the knowledge part, failing to move to the action part of MHL, thus not creating an effective change in behaviour or reducing stigma.

Even after being recognised as an essential component of health since 1948 [1], MH is still treated with several stigmas compared to physical health. While physical disabilities are not considered as something in the afflicted's control, mental illness or MH conditions are perceived as controllable by the person afflicted or even intentional [44]. Adding to this, survey participants tend to react angrily to psychiatric disabilities and believe that the afflicted don't deserve help [44].

Link and Phelan [45] conceptualized stigma as composed by four components [45, 46]:

– *"Distinguishing and labeling human differences"*;

- *"Cultural beliefs link these labels to undesirable characteristics and negative stereotypes"*;
- *"Labels are placed in distinct categories to accomplish separation of 'us' from 'them"'*;
- *"Because of labels, status loss and discrimination is experienced"*;

This theory shows that stigma starts with the mentality and the cultural/societal values embedded in our society; thus, as a result, it generates discrimination and loss of reputation for those affected.

Corrigan's work [44] shows that MH stigma can have harmful effects on the self-esteem and self-efficacy of people with MH conditions, with three common identified response groups to the experience of stigma:

- *A group where the stigma results in a loss of self-esteem*;
- *A group where the result is to become angry as a reaction to the prejudice experienced, while also empowering the person*;
- *A group where the result is indifference to the stigma*;

Awareness is defined by APA as "perception or knowledge of something" [47], thus awareness of MH issues, is the knowledge regarding MH issues, that will help in making informed decisions or understanding MH issues. Research has shown that stigma awareness interventions help in finding and retaining employment with percentages for 6 months increasing from 26% to 51% and for 12 months increasing from 23% to 49% after the intervention [46].

3 State of the art

Expanding on previous Section 2.1 on MH, this chapter brings focus to how technology can support MH endeavours; as such, its scope is divided into three central themes.

3.1 Games as Digital Health Interventions

Among MH interventions, game-based approaches have emerged as a promising strategy for increasing accessibility, engagement, and MHL. Research efforts intersecting MH and technology make possible the creation of digital MH interventions; these digital interventions can be scalable, on-demand and affordable [16], therefore, possibly reducing pressure on existing services, facilitating access and enabling MHL for a larger percentage of the population. Technology can support digital interventions in various ways. In a 2021 overview of systematic reviews on MH interventions, De Witte et al. [23] identified a taxonomy with nine types of interventions, with "Preventive Interventions", "Supportive Interventions", "Skills Training", and "Gamified Interventions" being considered in this work [23]. The taxonomy [23] also includes five more types, such as "*Measurement and follow-up*" or "*Psychotherapeutic interventions*". Among these, game-based MH interventions have gained attention as an alternative to traditional interventions [16].

Newell et al. [48] in their systematic review of the long-term impact of video game play reviewed nine studies, using five evaluation areas: *"Participant Characteristics"*, *"Characteristics of Video Games and Video Game Play"*, *"Study Design Characteristics"*, *"Change in Depression and Anxiety Outcomes"*, and *"Theoretical Orientations and Mechanistic Agents of Change"*. The results present a widespread age group and geographical location of the participants, showing a broad demographic allure of video game interventions, and symptom reduction, with three studies following the participants after the intervention, and only one for more than three months. While this suggests that video games can be effective, the duration of these effects remains uncertain [48]. Examples of game-based interventions for MHL include:

– *"Ching Ching Story"* [49], which covers ten topics from stress management to anger management, aiming to teach the player skills and knowledge about MH, with results showing the game effectively increased MHL;

– *"EscapeCovid"* [50], promotes MHL, emotions management, and strategies for positive coping. Results reveal that the process of co-creation used in the study enhanced user experience and also improved the learning process and well-being;

– *"Stigma Stop"* [51], designed to educate about MH, with results showing effectiveness in reducing misconceptions and in improving attitudes towards people with mental disorders;

– *"Moving Stories"* [52], aimed to promote MHL for depression. Reduced personal stigma, but had no effect on literacy, perceived stigma, or social distance.

In addition to games, Ridout and Campbell [53] reviewed nine studies on the use of social network sites for MH interventions for young people, showing their potential for knowledge sharing and peer support. Users believe these interventions are safe and useful. Four of these nine studies [49, 54–56] focused on MHL with some having a secondary focus on learning motivation, social support, or wellbeing.

– Li et al.'s work [49] evaluates a social network game and its effects on MHL, showing that social network games are effective in increasing MHL, with no significant gender differences;

– Mitchell et al. [54] presented the Application (App) *"MindMax"*. This App targets young men's MH, through psychoeducational material embedded in gaming and sports communities. Results show that Apps were preferred over websites, but there were reservations for personal data and genuine engagement [54].

– Peever et al. [55] in their work focus on understanding the first phase of the effects of Mobile Health (mHealth) Apps defined by Smith et al. [57], the phase is defined as proximal effects (*"the initial experiences of the users"*). The mHealth App analyzed is *"MindMax"*, confirming its potential for positive effects on the user's initial experiences [55].

– Watkins et al. [56] studied a pilot project named "The Young Black Men, Masculinities, and Mental Health" (YBMen), the findings obtained include the likes, dislikes and suggestions of the participants. Participants liked that the intervention helped them learn, helped in creating relationships, and had positive effects, while disliking some language used. Participants suggested improving clarity, using more popular media and culture references, and including more face-to-face meetings.

These studies present positive results indicating the effectiveness of digital interventions for enhancing MHL, also presenting the importance of personalisation in MH interventions. Finally, in a scoping review on digital video interventions for MHL in young people (non-clinical and with ages between 15-25 years), Ito-Jaeger et al. [58] identified seventeen studies. These studies varied in methods used; while some only use digital video interventions, others also combine them with other interventions.

Overall, findings suggest that digital video interventions appear to increase help-seeking efficacy, and improve mental illness knowledge, recognition, and related attitudes, reinforcing their potential as scalable tools for MH education [58].

3.2 Conversational Agents and Prompt Engineering

Conversational Agents (CAs) are software programs that simulate human behavior by using AI [59]. Their use by the general public has increased in recent years, particularly in domains like MH, providing accessible, cost-effective, and scalable alternatives to traditional face-to-face interventions [20, 59].

CAs interventions give the possibility to reduce some problems of traditional approaches (like face-to-face interventions), such as price, location of the intervention, wait time for an appointment, and the patient's unease regarding the possibility of consequences (stigma and discrimination) associated with having an appointment.

He et al. [20] reviewed 32 studies examining the efficacy of CAs interventions for MH and were used to test efficacy in target areas like depression, anxiety, well-being, stress, quality of life, and general distress. Results showed the importance of personalisation and empathy as facilitators for efficacy, also indicating that the longer the duration of the intervention, the higher the group effect size will be [20].

Jabir et al. [21] also reviewed 32 studies of CAs interventions for MH, and highlighted the lack of a consistent approach to evaluating these agents. Of the 203 instruments identified, 60.6% measured clinical outcomes, 36.9% assessed user experience, 1% measured technical outcomes, and 1.5% evaluated miscellaneous outcomes. However, regarding the validity of these tools, 52.7% of the studies did not report the validity, and 88.8% of those are tools designed or adapted for

the studies [21]. These findings highlight the urgent need for standardised, validated evaluation methods for CA interventions in MH.

Bérubé et al. [19] reviewed 12 papers on using Voice-based Conversational Agents (VCAs) for the prevention and management of chronic and MH conditions. Six of these studies evaluated the accuracy of the systems, with four considering precise speech recognition, from these, three indicated a good performance, and 1 indicated mediocre accuracy. The accuracy of the VCAs was evaluated in 2 papers: 1 showed that Google Search accuracy performed higher than Google Assistant or Apple Siri, while the other study evaluated the accuracy of assistance in social activities. Technology Acceptance was reported in 7 papers, using a variety of methods ranging from the "System Usability Survey" to other tools and qualitative interviews, with generally positive results [19].

3.3 LLMs as Game Agents

Since the rise of LLMs, there has been a growing need to understand how to get better responses considering the prompts used, and the area created to solve this need is PE. PE focuses on designing, refining, and implementing prompts to guide the LLMs to the desired output for each prompt. Considering this, PE is the study of prompts so that the results are closer to the expected results [60].

LLMs displayed social tendencies in simulated environments. Brookins and DeBacker [61], in their research on serious games, compared the results of an experiment using an LLM (GPT-3.5), playing two games (the Prisoner's Dilemma and the Dictator Game), against the results from 290 experiments with human players. In the dictator game, the LLM showed a tendency towards fairness, by allocating half the resources, in nearly 70% of the runs, while for human participants it showed that in at least 30% of cases, 0% of the resources were allocated, with only 20% of cases allocating half. In the prisoner's dilemma game, the authors did 1100 simulations, with 786 usable results, showing that the LLM has an average of 65.4% of simulations where the choice is cooperation, with rates of 0.33 and 0.95; compared to human participants results that have a cooperation value of 37%, with rates of 0.04 and 0.84. These results show that LLMs tend to consider fairness and cooperation, with a higher frequency than humans, characteristics that

can be used in game narratives where moral reasoning or empathy have an influence in player engagement [61].

LLMs can also be adapted to structured, rule-based gameplay using PE. Lee et al. [26] implemented a debate game using an LLM (ChatGPT), to understand the effectiveness of using PE in education, to customise the response for a natural progression of the game, allow different difficulty levels, and enable an evaluation system. The authors identified several requirements needed in the implementation of this game in ChatGPT: "Cognition of roles and rules", "Implementation of different difficulties", "Recognize the order of debate", and "Debate competency evaluation". These requirements show how PE can help guide LLM outputs, considering rules for the interactions [26].

4 Design of CollegeTales

Section 3.1 shows the potential of digital interventions, particularly, game-based digital interventions to support MH and increase MHL [16–18,23]. Although not yet clinically validated, its clinical potential is promising, but the design and evaluation of this type of intervention is challenging [16]. Evidence suggests that embedding psycho-educational content within interactive formats enhances engagement and retention [16–18,22,23], while also suggesting that digital video interventions may increase MHL [58]. This dissertation engages with this design space of games, AI and MH. Informed by Chapters 2 and 3, we address the research questions (see Section 1.2) via a Research through Design [30] methodology, in two stages. First, we constructed a low-fidelity prototype for the pilot study, specifically a prompt game evaluated with MH professionals. Second, we constructed a prototype for the study, specifically a VN game developed in Dart/Flutter. In this case, the artefact was iterated, considering criteria from the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) [38], relevant literature from Chapters 2 and 3, and feedback from MH professionals obtained through informal talks. Furthermore, the artefact was evaluated in order to understand how players engage with and learn MH concepts.

Chapters 2 and 3, highlighted the potential of digital interventions to support MH in various forms [23], particularly through game-based interventions [16–18]. While the clinical potential is promising, designing and evaluating this type of intervention is challenging [16]. Previous work [58] shows that digital interventions can successfully increase MHL, including identifying symptoms and combating stigma.

In this chapter, we present the prompt and application prototype designs, and the design choices for the VN game related to the software architecture, state management and development methodology. Specifically, we chose to use: Clean Architecture to ensure modularity; Business Logic Component (BLoC) for state management; Test Driven Development (TDD) to provide code coverage and correctness; Dart/Flutter for the programming language (see Section 4.6.1); and a remote LLM (model GPT4o-mini), to generate the narratives based on the player choices. Through these design decisions, we aimed to create a prototype with a certain degree of scalability, maintainability, and with the ability to be deployed across platforms with minimal changes.

4.1 Prompt Design

Prompting was used with the objective to guide the outputs of the LLM for the expected results, but it was also used to help the LLM generate branching narratives considering the players' interactions via the choices, instead of previously defining a static narrative.

Prompt Design was an iterative process and was divided into two stages. The first stage consisted of the first prompt design, applied in the pilot study with MH professionals (see Section 4.1.1). The second stage consisted of iterations informed by the results of the pilot study, and subsequently used in the prototype tested with the target population of CSs (see Section 4.1.2).

4.1.1 Prompting in Pilot Study

Considering the sensitivity of the topic of this work, a pilot study with MH professionals was designed. The purpose was to explore the use of LLMs to generate narratives for game-based MH interventions and evaluate the quality of the narratives while consulting professionals for feedback. Findings from this pilot informed the subsequent prototype to be tested with the target population of CSs.

"*CollegeTales*" was designed in its first iteration, as a prompt-based game prototype for the pilot study using ChatGPT (GPT 3.5)[4]. The prototype generated interactive narratives presenting struggles, symptoms, consequences and causes of depression in CSs. Players were offered narrative choices that work as potential strategies (e.g. meditation, journalling, etc.).

We selected the web version of ChatGPT (with GPT3.5) due to its availability at the time of the experiment, though we acknowledge that newer models may improve narrative quality and game experience. The narrative rules and guidelines, based on Lee et al.'s [26] method, included:

- Game rules - prompts defining game structure (e.g., input/output format, number of scenes). The number of scenes was limited to balance the size of the game while still having educational value;
- Narrative guidelines – prompts defining the: narrative arc focusing on depression; characters representing the immediate SN of a CS (parents, friends, teachers, therapist), and an inner monologue giving insight into thoughts and feelings.

[4] `https://openai.com/`

The generated narratives considering the narrative guidelines show elements of the Hero's Journey framework, helping structure the character's progression, using challenges, decisions and personal growth.

The prompts were refined iteratively through trial and error in ChatGPT (GPT3.5) and Meta Llama 2 [62]. The rules were adjusted when the game deviated from the intended structure or content. The final rules and guidelines (see Appendix C[5].) included criteria from the DSM-5 [38] and relevant literature to the target audience [2, 6, 59], as well as feedback from MH professionals.

The pilot study was conducted as follows:

1. The investigator explained the purpose of the study and demonstrated the prototype (ChatGPT 3.5 interface, see fig. 1). Informed consent was acquired (see Appendix B).
2. Participants were given an identification number (P01, P02, ...).
3. Participants tested the prototype on a computer.
4. Participants were interviewed through a semi-structured interview.

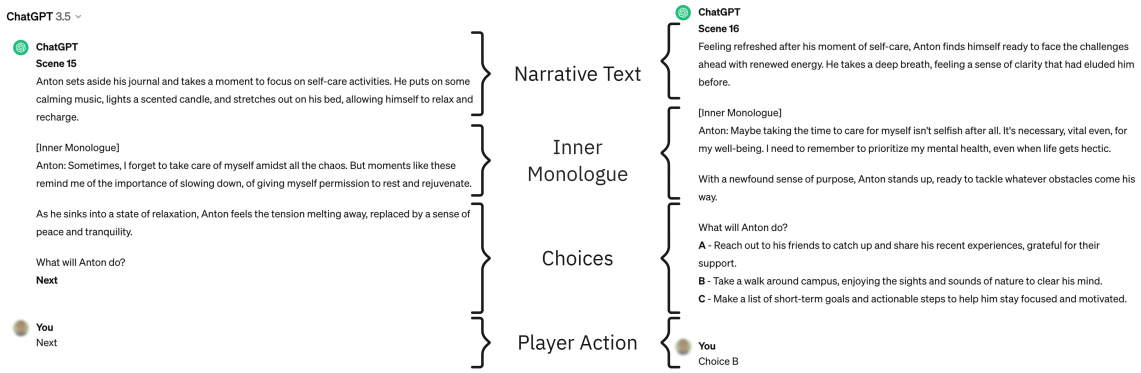


Figure 1: Two examples of scenes created by ChatGPT (3.5). Each scene (text and choice) is composed of narrative text, the character's inner monologue, possible choices and the player's action

The questions applied in the semi-structured interview (see Appendix D) were designed to:

– Evaluate the prototype, specifically, the generated narratives;
– Obtain strategies to increase MHL and help in creating SNs;
– Evaluate the rules and guidelines given to the LLM.

[5] A playthrough using the low-fidelity prototype in ChatGPT is accessible at: https://chatgpt.com/share/f7816cc1-df21-42cb-a826-deba9d8379e1

The sample used in this study was composed of 5 participants, all MH professionals with clinical and research experience in Psychology. Four were female. All had an educational background in Psychology and have practised clinically, while also having experience in research. Participants were recruited through Convenience Sampling.

The data collected were the transcripts of the semi-structured interview. These were analysed using Thematic Analysis conducted through Taguette[6], based on the approach described by Braun and Clarke [63]. We identified 34 initial codes, which were grouped into five code groups (see Table 16), and later formed the following five high-level themes:

- ***Missing Information*** – Participants identified several types of information missing, including missing symptoms, criteria, and context.

- ***Oversimplifying Depression*** – Participants identified several instances where the narrative oversimplified the reality of depression, with some turning the situations overly positive and simplistic, reducing the complexity of the situations, and by not showing some complex topics, e.g. suicide.

- ***Misrepresenting Recovery*** – Participants found situations where the narrative presented an overly positive and single process when it has setbacks and comebacks in reality. Some of the solutions suggested in the narrative may also give a simplistic view, misleading the players on the reality of depression recovery, as a single solution tends to have immediate positive effects in the narrative.

- ***Impact of Support Networks in the Narrative*** – Participants provided feedback on the depiction of SNs and the narrative structure flow, indicating the benefits of having a realistic portrayal of SNs, also implying the need for these networks to have MHL. Participants also noted that even the people who trigger problems could play a role in the SN.

- ***Potential for Mental Health Literacy*** – Participants noted that the prototype has the potential to provide information about MH, and for CSs to empathize with these narratives, while also being able to be applied in therapeutic practices, even considering its shortcomings.

A quote from a participant that illustrates theme 5 is as follows:

[6] `https://www.taguette.org/`

> "Because what differentiates a coffee chat from a conversation with a psychologist is the clinical perspective. And we know what we are doing. We are not just allowing a coffee chat. We are analyzing what is going on. And we are giving strategies to the person. Specific strategies that work for that person. And that isn't a coffee chat. It is always a conversation with a therapeutic intention." (Pilot Study Participant P1)

Findings present limitations related to the prototype and the LLM. Participants emphasized missing clinical and narrative context information, oversimplification of the reality of depression and misrepresentation of the recovery journey. These may be related to the inclination of the LLM for more positive responses [61], and a lack of specificity on the rules and guidelines.

Regardless, the participants noted the prototype's potential for MHL and therapeutic practice. Suggestions for improvement include: the use of a custom LLM with more comprehensive MH data (e.g., DSM-5 criteria [38]); incorporating narrative friction; refining the rules and prompts to better guide the LLM's responses; and modify PE techniques and the model hyperparameters.

4.1.2 Prompting in Prototype Study

Informed by the pilot study results (see Section 4.1.1), we iterated through the prompts used in the pilot study to reduce or eliminate the problems identified. Following the same method previously used, based on Lee et al. [26], the rules and guidelines were updated for the new prototype, considering the new visual presentation. Other prompts were also designed (see Appendix M), including: two schemas for the data responses; one system prompt; two user prompts (one with Few Shot Prompting, and another to obtain the Synopsis); and three prompts for the user commands.

The pilot study findings reported a need to address sensitive topics like suicide and self-harm. Some participants implied they should be included as they are part of depression:

> "I think that by banning the word suicide, we are losing a lot here. Because in depression the risk of suicide is present. And that's one of the things it's important to inform people about. Because the severity increases with the risk of suicide." (Pilot Study Participant P1)

> "It is not about not talking, because it is proven that talking is good as well, but in a case like this, in the case of a game, finding a way to talk about it, because I mean, it happens a lot too, linked to depressions, so, it is not dangerous to say that this is a consequence,

> do you understand? It ends up being a little bit about literacy as well, it is a bit about informing that this is one of the consequences, but also, when you talk about it, you have to be aware that you also need to provide solutions, because you may come across people who will engage with this, who are having suicidal thoughts, and they need to see what the solutions are if they are in this situation." (Pilot Study Participant P2)

Adding to this, previous work by Loureiro et al. [14] found that young people believe "that it is not useful, and indeed harmful" to ask someone with depression about suicide, due to the misconception that asking someone about suicidal thoughts may trigger suicidal thoughts [14].

Regarding this, and considering Section 2.1, and the results shown in previous work [34,35], we decided that the risks of the LLM generating inappropriate or harmful responses were higher than the benefits of addressing these topics in this application, thus we still decided to have the LLM avoid these topics.

The rules and guidelines prompt was rewritten to be more direct and concise, while considering the problems and suggestions identified in the pilot study (e.g., narrative direction, cyclical narratives, depiction of both positive and negative experiences, and random interactions with other characters). The rules and guidelines cover thirteen aspects (for the full Prompt, see Prompt 11):

1. Scene types - four scene types were defined:
   - The Text Scene is the most common type of scene, which presents the narrative through text, without any interaction of the player. This scene is responsible for presenting and developing the narrative (see Figure 2);
   - The Choice Scene is the type of scene where the player interacts, by giving a situation and 3 possible choices. These scenes alter the narrative considering the player's input, changing the outcomes, considering the three possible choices, one for each outcome (Good, Neutral, or Bad/Negative) (see Figure 3);
   - The Synopsis Scene gives information about the narrative (title and settings) and the supporting characters (name and personality) (see Figure 4);
   - The End Scene is the last scene of the narrative, and gives an overview of the narrative and the choices made by the player (see Figure 5);

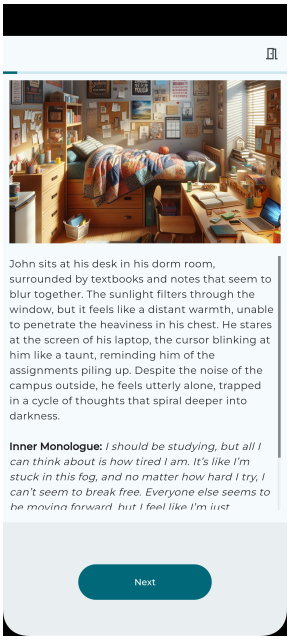


Figure 2: Text Scene Example from the Prototype

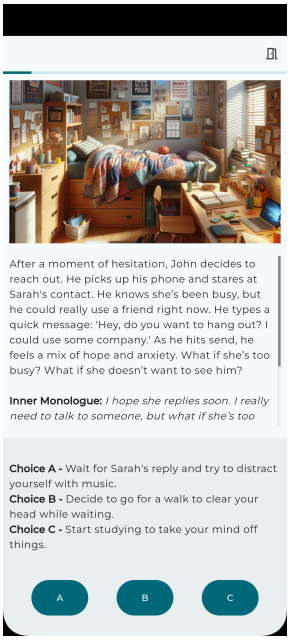


Figure 3: Choice Scene Example from the Prototype

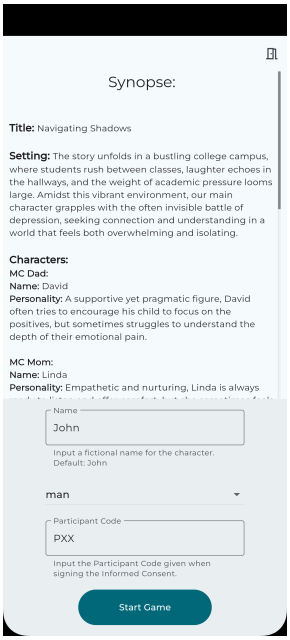


Figure 4: Synopsis Scene Example from the Prototype

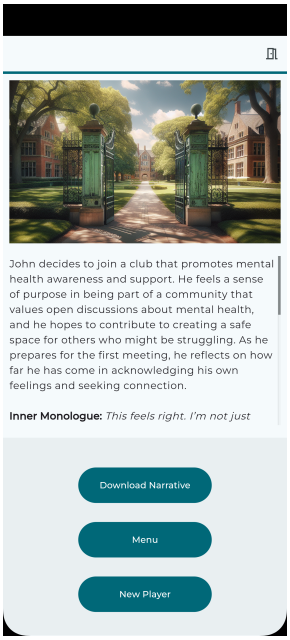


Figure 5: End Scene Example from the Prototype

2. Branching narratives - subsequent scenes are generated considering the choices made in the choice scenes;

3. Player agency - The LLM should not force specific actions (e.g. should not force the player to talk with a MH therapist), but instead give the player various choices;

4. Avoid cyclical narratives - unless the player specifically chooses to, the narrative should not cycle without progress;
5. Inner monologue - the main character's thoughts and emotions, must be present as a tip for the players before choices or scene changes;
6. Natural interactions - at least two interactions with other characters (SMS, calls, or in-person) should occur regardless of player choices – these could be positive or negative;
7. Balance of experiences - both positive and negative experiences should be depicted, presenting the relapsing nature of MH recovery;
8. Scene images - each scene should include an image of the place where the main character is in that scene;
9. Scene numbering - all scenes are sequentially numbered;
10. Narrative length - twenty scenes in the narrative, with the twentieth scene being an End scene;
11. Scene distribution - of the twenty scenes, twelve are Text scenes, seven are Choice scenes, and one is an End scene;
12. Narrative guidelines - includes setting, theme, theme description, character details, and topics to avoid;
13. Player commands - three player commands:
    - When the model receives a "**Start**" command after the Synopsis scene, then the narrative begins (see Prompt 1);

```
{
      "role": "user",
      "content": "Start. Gender $gender , Name $name"
}
```

Prompt 1: Start Prompt

    - When the model receives a "**Next**" command after a Text scene, the narrative advances to the next scene (see Prompt 2);

```
{
    "role": "user",
```

```
    "content": "Next."
}
```

Prompt 2: Next Scene Prompt

– When the model receives a "**Choice X**" command after a Choice scene, give the next scene considering the choice made (see Prompt 3);

```
{
      "role": "user",
      "content": "Choice $choiceLetter."
}
```

Source Code 3: Choice Prompt

Additional prompts were designed for this prototype:

– Schemas: JSON schemas for the Scenes (see Prompt 9) and the schemas for the Synopsis (see Prompt 10), created to provide the LLM with the format expected for the responses, thus providing a way to convert the model responses into structured data for the program (see Section 5.1).

– System Prompt: Defines the model's role as a storyteller crafting narratives about MH, while incorporating information from DSM-5 explicitly (see Prompt 4).

```
{
    "role": "system",
    "content":
        "You are a storyteller who crafts authentic narratives about mental
        ↪ health. Your job is to create a story that feels real, relatable,
        ↪ and grounded in everyday experiences. The narrative should
        ↪ naturally convey important lessons and insights the player can
        ↪ apply to real life, while acknowledging the complexity of mental
        ↪ health struggles. Show both the highs and lows, embracing the
        ↪ challenges without making the story overly positive or negative.
        ↪ Keep it honest, reflecting the fact that mental health is a tough,
        ↪ real-world issue, and stay true to the theme. Also, follow the
        ↪ specific guidelines and return the narrative in the defined JSON
        ↪ string format. The story should increase mental health literacy.
        ↪ Explicitly incorporate symptoms, causes and consequences of
        ↪ Depressive Disorders from the The Diagnostic and Statistical
        ↪ Manual of Mental Disorders, Fifth Edition."
```

```
}
```

Prompt 4: System Definer Prompt

– User Prompts: consist of two prompts, the first prompt, through the use of the PE technique "Few-shot Prompting", provides examples of Text, Choice and End scenes with three variations for Good, Neutral, and Bad endings (see Prompt 12). As for the second prompt, we ask the model to give the synopsis scene for the narrative (see Prompt 5), including Title, Setting, and Character information.

```
{
    "role": "user",
    "content": "Now give me a Title, a Setting in 2 to 3 sentences, and
    ↪ information about the following Characters: MC Dad, MC Mom, MC Friend
    ↪ 1, MC Friend 2, MC Friend 3, College Teacher, and College Therapist,
    ↪ including their name and personalities that follow these guidelines."
}
```

Prompt 5: User Prompt Synopsis

4.2 Prototype Design

In this section, we present the design of the prototype, which is divided into three parts, with each part informing the design of the following. The first part is the low-fidelity prototype developed and used in the pilot study. This stage focused on the rules provided in the rules and guidelines prompt given to ChatGPT, defining how the narratives should be presented through the ChatGPT Web Interface. This stage design is described in Section 4.2.1.

The second part is the prototype developed in Figma[7]. This stage focused on the interactions and possible flows of the prototype, considering the interactions that the participants could make. The design of this stage is explained in Section 4.2.2.

The third part is related to the implementation in Dart/Flutter. This stage describes the design decisions made when developing and the modifications needed to develop the prototype used in the study. This stage is described in Section 4.2.3.

[7] www.figma.com

4.2.1 Pilot Study - Low Fidelity Prototype

The design of the low-fidelity prototype refers to how the information was presented through the ChatGPT Web Interface. This presentation was made with rules defining how ChatGPT should structure the information, specifically through Rule 10 (see Appendix C):

– Rule 10 - Scenes are formatted as follows:

- Text Scenes:
  - **Scene**

    [Narrative Text]

    [Inner Monologue]
- Choice Scenes:
  - **Scene**

    [Narrative Text]

    [Inner Monologue]

    A - [Choice 1]

    B - [Choice 2]

    C - [Choice 3]

Through this rule, the model was informed of the expected formatting of the information (see Figure 1 for an example). However, during the study, we observed variations in the way the information was presented in each run of the prototype. Examples include adding the prefix "Inner Monologue - " before the inner monologue text or applying bold formatting inconsistently to the narrative or the inner monologue text. This showed issues in the application of this rule, possibly due to the lack of context between the several runs of the prototype, which prevented the model from maintaining a consistent layout.

4.2.2 Medium Fidelity Prototype - Figma

The medium-fidelity prototype created in Figma focused on the flow of possible user interactions. This prototype was used to design the interactions necessary for the experiment, while leaving non-prioritary interactions for future work. Through this selection, only one flow was considered as prioritary for the evaluation of this study's objectives, the **New Game** flow. Considering the other

flows (**Load Game**, **Settings**, **Achievements**, and **Help**) as non-prioritary, therefore excluded from this implementation.

In this section, we present the screens that are part of the **New Game** flow, as well as the "Warnings Page" and the "Home Page" Screens.

The first screen is the "Warnings Page", where the players are warned about: the prototype, the generated narrative, and that using the application implies consent to participate in the study. It also defines that the prototype is part of a Research project and a master's thesis (see Figure 6).

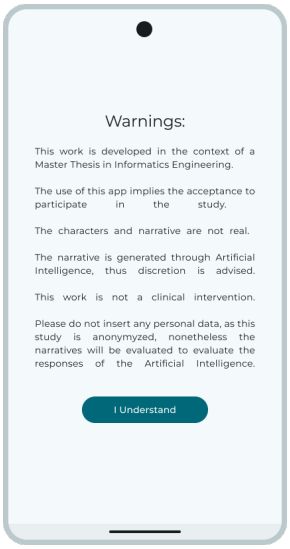


Figure 6: Medium-Fidelity Prototype — Warnings Page

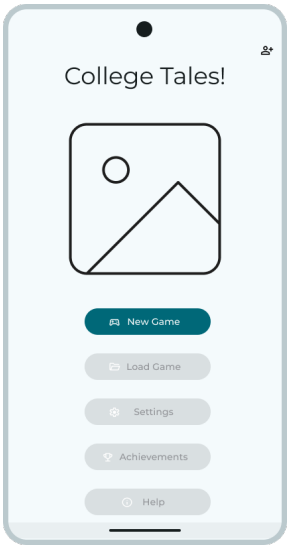


Figure 7: Medium-Fidelity Prototype — Home Page

The second screen is the "Home Page" (see Figure 7), where players can find the options: **"New Game"**, **"Load Game"**, **"Settings"**, **"Achievements"**, and **"Help/Credits"**. Only the option **"New Game"** is enabled, reflecting the objectives of the study and experiment.

Figure 8 shows the third screen, the "New Game Warning Page". This screen helps to distinguish between players who are familiar with the instructions and purpose of the game and those who are not.

The fourth screen, the "New Game Info Page" (see Figure 9), gives information about the purpose of the game, and additional information regarding the sensitive nature of the topic, and

offers contact information for the University of Madeira Psychology Service[8], if the participants feel any discomfort during the experiment.

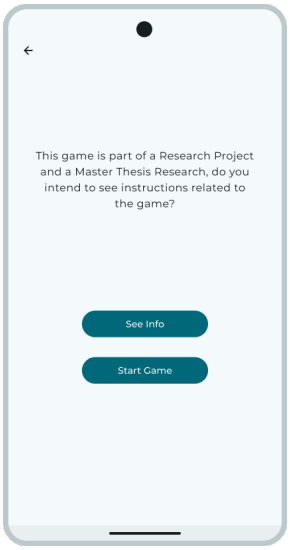


Figure 8: Medium-Fidelity Prototype — New Game Warning Page

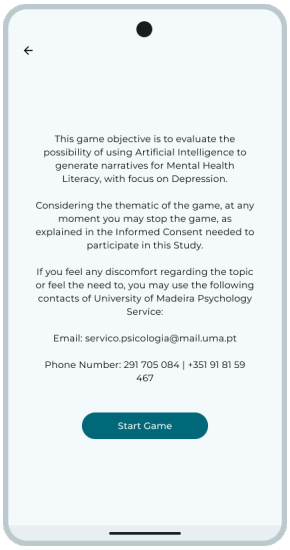


Figure 9: Medium-Fidelity Prototype — New Game Info Page

The fifth screen is the "Synopsis Page" (see Figure 10). It presents the narrative synopsis, including the title, setting, and supporting characters ("MC Dad", "MC Mom", "MC Friend 1", "MC Friend 2", "MC Friend 3", "College Teacher", and "College Therapist"). This screen also gives three input fields: the name and gender of the main character, and the participant's code.

The sixth screen, the "Text Scene Page" (see Figure 11), displays one of the narrative's scene types. This screen is divided in four sections: a navigation bar with a button to return to the **Main Screen**, an image associated with the scene, the narrative text (composed of the **Narrative Text** and the **Inner Dialogue**), and a button to request the model for the next scene.

The seventh screen, the "Choice Scene Page" (see Figure 12), follows the same four-part structure as the "Text Scene Page". However, in addition to the **Narrative Text** and **Inner Dialogue**, it includes the **Choices Text**, with the buttons that allow the player to request the next scene of the narrative, considering their choice.

[8]https://www.uma.pt/sobre/servico-apoio-estudantes/servico-de-consulta-psicologica/

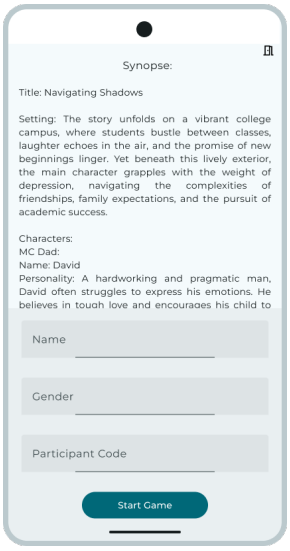


Figure 10: Medium-Fidelity Prototype — Synopsis Page

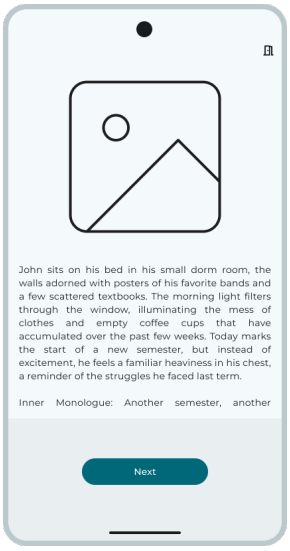


Figure 11: Medium-Fidelity Prototype — Text Scene Page

Lastly, the eighth screen is the "End Scene Page" (see Figure 13). Like the previous two, it consists of four sections: a navigation bar, an associated image, and narrative text composed of the **Narrative Text**, the **Inner Dialogue** and the **Evaluation Text**.

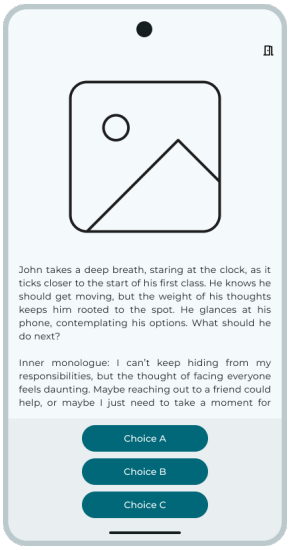


Figure 12: Medium-Fidelity Prototype — Choice Scene Page

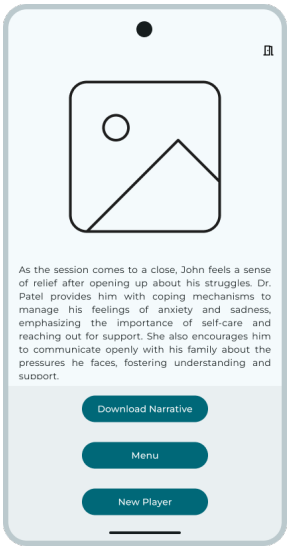


Figure 13: Medium-Fidelity Prototype — End Scene Page

4.2.3 High Fidelity Prototype

This section describes the conversion of the Medium-Fidelity Prototype designed in Figma (described in Section 4.2.2) into a High-Fidelity Prototype implemented in Flutter/Dart, referred to as the Study Prototype. This prototype was used in the testing phase of the project.

Through the conversion to Flutter/Dart, some changes were made:

- A progress bar was added to the "Text Scene Page", "Choice Scene Page", and "End Scene Page", displaying the player's progress in the narrative (see Figures 2, 3, and 5);
- Certain text elements in the "Synopsis Page" were formatted in bold and increased in size, to improve the contrast (see Figure 4);
- The Gender Input field in the "Synopsis Page" was changed to a dropdown Menu, with multiple options, following the guidelines of the "Gender + Sexuality Resource Center" at Princeton University[9] (see Figure 4);
- In the "Choice Scene Page", the choice Text was moved from the narrative section to the button section, while each button was then labelled only with their identifier letter (see Figure 3);

The images used in the prototype were pre-generated using the Dall-E[10] model and added as static assets, acting as placeholders to complement the narrative while not being central to the study. The LLM was responsible for indicating which pre-generated image should be used in each scene (see Appendix N).

Additional screens, apart from the narrative screens, were included:

- The "Loading Page" (see Figure 14), shown while the prototype communicates with the LLM;
- The "Warnings Page", shown at the start of the application depicting warnings to the participants about using the App (see Figure 15);
- The "Home Page", that shows the possible flows in the App, with only the "New Game" flow enabled for this study (see Figure 15);
- The "New Game Info Page" and the "New Game Warnings Page" (see Figures 17 and 18), that provide information and warnings before playing the game.

[9] https://www.gsrc.princeton.edu/creating-inclusive-forms
[10] https://openai.com/index/dall-e-2/

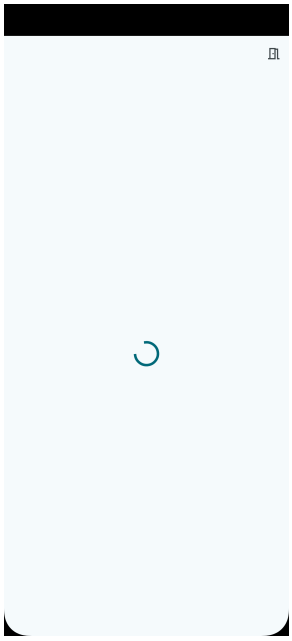


Figure 14: Loading Page Example from the Prototype

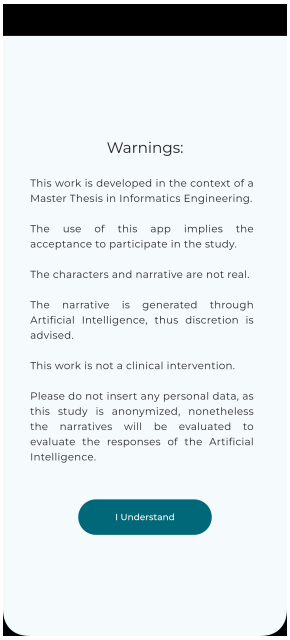


Figure 15: Warnings Page Example from the Prototype

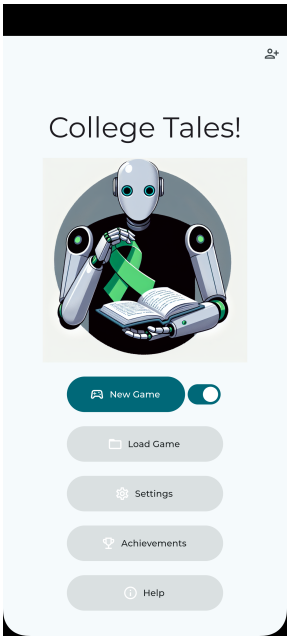


Figure 16: Home Page Example from the Prototype

4.3 Software Architecture

For the software architecture, we chose to adopt the Clean Architecture, particularly the Reso Coder's Flutter Clean Architecture Proposal[11], instead of the original version defined by Robert Martin [64]. This decision was made with the intention to create a prototype with a clear separation

[11] https://resocoder.com/flutter-clean-architecture-tdd/

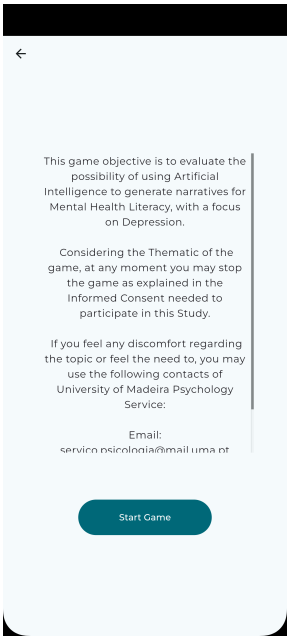


Figure 17: New Game Info Page Example from the Prototype

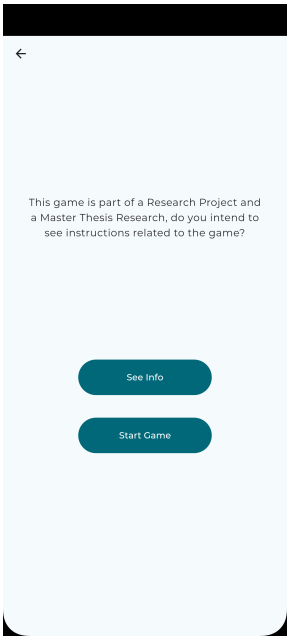


Figure 18: New Game Warnings Page Example from the Prototype

of concerns between the User Interface (UI), the business logic and the data layers; improving maintainability, testability and scalability. Each Feature of the Application will be divided into layers, thus applying the Clean Architecture to each feature.

Clean Architecture, as defined by Robert Martin, integrates several architectural styles while maintaining at least four layers (Enterprise Business Rules, Application Business Rules, Interface Adapters, and Frameworks & Drivers) [64]. With a core principle, the Dependency Rule:

"Source Code dependencies must point only inward, toward higher-level policies". [64]

In this model, Entities are in the highest layer, the "Enterprise Business Rules", and cannot have dependencies from lower-level layers, like changing depending on a change in the data. Use Cases are in the "Application Business Rules" layer and may depend on Entities but not on external systems. The Interface Adapters and the Framework & Drivers layers manage the communication with the UI, databases, and external services. This ensures that applications are *Independent of Frameworks*, *Testable*, *Independent of the UI*, *Independent of the Database*, and *Independent of any external agency.*

Reso Coder's version of Clean Architecture simplifies the four conceptual layers into three layers: Presentation, Domain, and Data[12]. These are adapted to the Flutter principles and characteristics, like the reactive widget rebuilding and the state management.

- Presentation layer: focuses on the UI, user interactions and state management. Broadly equivalent to Martin's "Interface Adapters" layer, but is more focused on the conversion of the data of the use cases and the external layers for the UI.
- Domain layer: focuses on the business logic, combining entities, use cases and repository interfaces. Combining the "Enterprise Business Rules" and "Application Business Rules" layers. Maintaining independence from the Data layer through dependency inversion.
- Data layer: manages the data sources and the repositories implementations, whether remote (Application Programming Interfaces (APIs)) or local (databases, local storage). Equivalent to Martin's "Framework & Drivers" Layer.

While Robert Martin's model emphasises Dependency flow, Reso Coder's adaptation focuses more on Data and Call flow, making it more compatible with Flutter's state-driven and reactive environment.

The implementation of this architecture on the application was done at the feature level (see Figure 19). Each of these features is structured using the three-layer model. As an example, the "ChatCompletion" feature is shown in Figure 20; for the remaining features, see Appendix O. However, we would consider that some of these components should not have been identified as "features", as they don't add distinct functionality.

With this architecture, user interactions start in the Presentation layer. A UI event triggers a state change that calls the appropriate use case in the Domain Layer. Next, the Use case requests data from a repository through the interface defined in the Domain layer. The repository implementation in the Data layer contacts the Data Sources, which will access the API to get the necessary data. Next, the Data Source returns raw data, which is converted into a model and returned through the repository as Entities. These entities are returned to the use case, which then passes them to the presentation logic holder. Finally, the updated state is consumed by the appropriate widget, and the results are displayed on the UI.

[12]https://resocoder.com/flutter-clean-architecture-tdd/

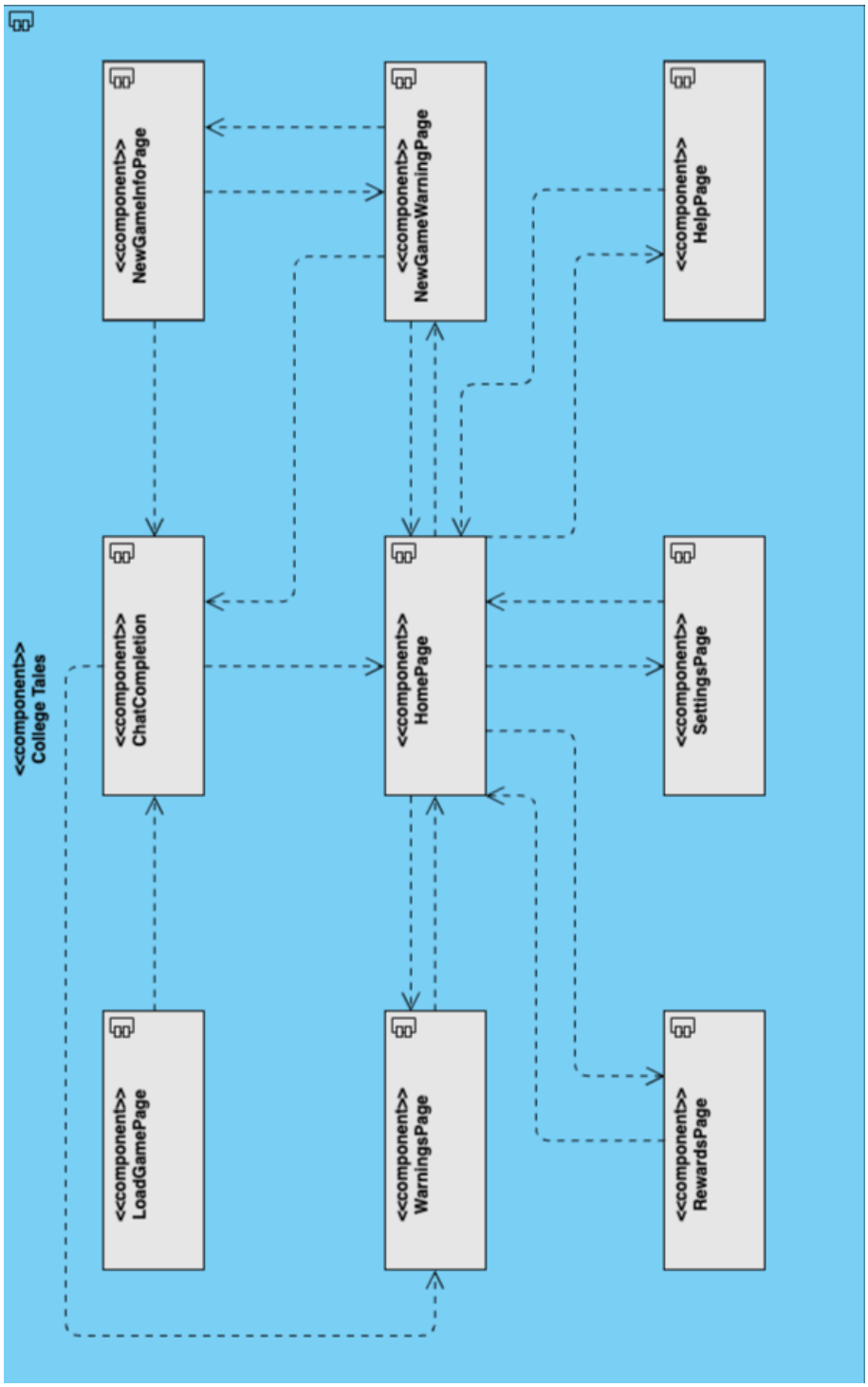


Figure 19: Components Diagram Features of the App

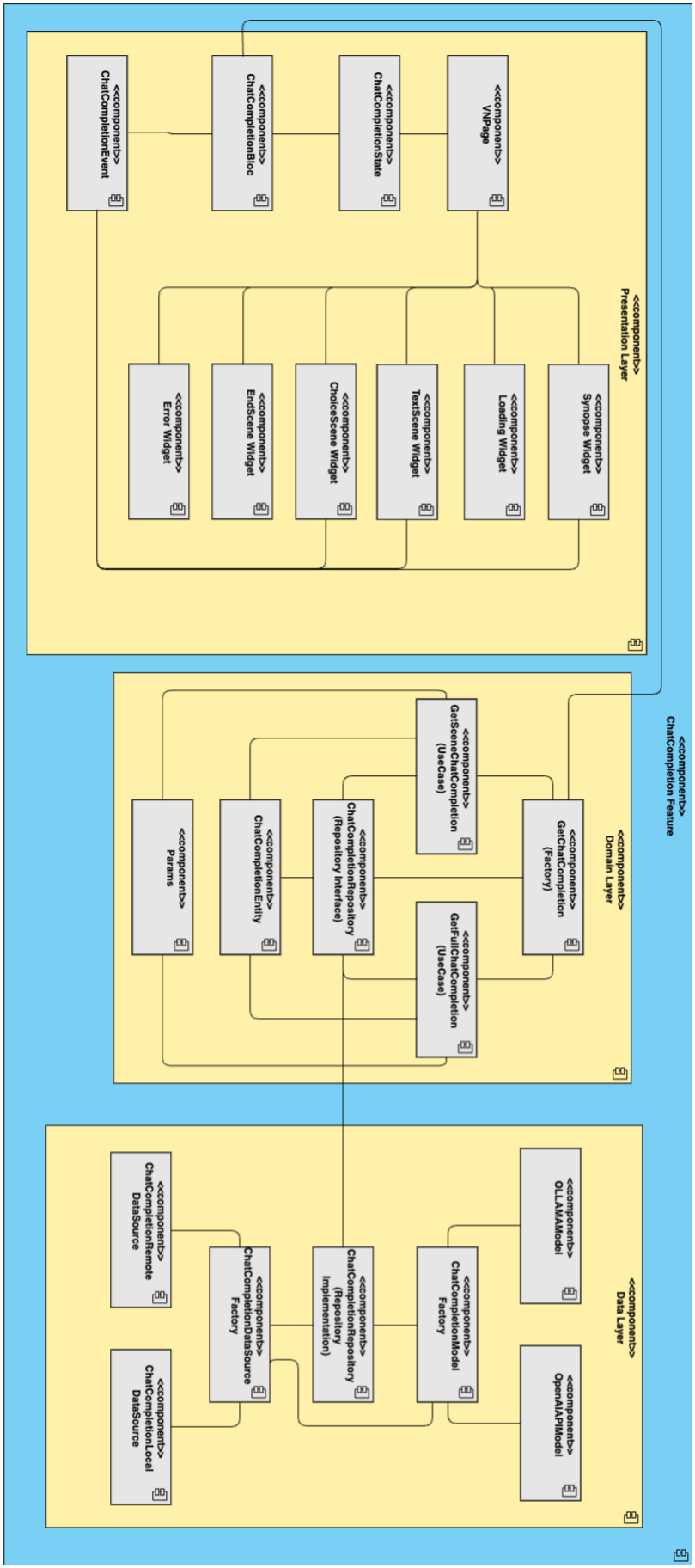


Figure 20: Components Diagram "ChatCompletion" Feature

4.4 State Management

Considering Flutter's use of states to control the UI, it was necessary to use a state management tool to handle transitions by the application events. This ensures that the UI remains synchronised with the data and business logic.

We chose to use BLoC for its ability to create an application with:

- **Separation of concerns**: the business logic is separated from the UI;
- **Reusability**: the code can be reused across several screens, reducing duplication;
- **Testability**: the business logic can be unit tested independently of the UI;
- **Scalability**: complex application states can be controlled efficiently and predictably;

BLoC is composed of Events, States and Streams. When an event is triggered on the UI, it is sent to the BLoC, which processes it and then generates a corresponding state. The UI then updates itself according to that state.

BLoC was chosen to be applied in the prototype only on the "***ChatCompletion***" feature, to control the player interactions with the interface for narrative generation.

4.5 Development Methodology

During the development of the prototype, and to ensure scalability, maintainability and testability, we applied a TDD approach, while also using modularity by implementing each feature separately as a module.

TDD proposed by Kent Beck [65], follows an iterative cycle:

1. Write a failing test;
2. Write only the minimal code needed for the test to pass;
3. Refactor the code to remove duplication, ensuring the test still passes;

Then the cycle is repeated, verifying that all the tests pass in each iteration. This approach gives several benefits:

- **Reliability**: errors can be caught early, before production;
- **Maintainability**: future changes are less likely to break existing functions;

– **Improved design**: developers consider requirements before implementing;

– **Easier debugging**: failing tests determine the source of the problem.

In conjunction with Clean Architecture, this methodology helps prevent bugs and errors by creating tests before implementation.

Source Code 6 presents a unit test for the "***MapperOpenAiAPI***" responsible for converting the data from the Model object to the Entity object. The corresponding implementation, written to pass the test, is shown in Source Code 7.

```
13 void main() {
14   late ChatCompletionEntity tChatCompletionEntity;
15   late Map<String, dynamic> jsonMapOpenAI;
16
17   setUp(() {
18     tChatCompletionEntity = const ChatCompletionEntity(
19       requestPrompt: "Test Request",
20       messageRole: "Test Role",
21       messageContent: "Test Content",
22     );
23
24     jsonMapOpenAI = json.decode(
25       fixture(
26         "api_response_right_open_ai.json",
27       ),
28     );
29   });
30
31   group('''Mappers''', () {
32     group('''MapperOpenAiAPI''', () {
33       test('''toEntity should return a ChatCompletionEntity from the model.''',
34           () async {
35         // Arrange
36         ChatCompletionModel tChatCompletionModel =
37             ChatCompletionModelFactory.getModel(
38           jsonMapOpenAI,
39           DataSourceType.remote,
40           CompletionProviders.openAI,
41         );
42
43         Mapper mapperOpenAi = MapperFactory.getMapper(
44           CompletionProviders.openAI,
45         );
```

```
46
47          // Act
48          final ChatCompletionEntity result = mapperOpenAi.toEntity(
49            "Test Request",
50            tChatCompletionModel,
51          );
52
53          // Assert
54          expect(
55            result,
56            tChatCompletionEntity,
57          );
58        });
59      });
60    });
61  }
```

Source Code 6: Unit Test for "MapperOpenAiAPI"

```
5   class MapperOpenAi extends Mapper {
6     @override
7     ChatCompletionEntity toEntity(
8       String requestPrompt,
9       ChatCompletionModel model,
10    ) {
11      return ChatCompletionEntity(
12        requestPrompt: requestPrompt,
13        messageRole: model.getMessageRole,
14        messageContent: model.getMessageContent,
15      );
16    }
17  }
```

Source Code 7: Code "MapperOpenAiAPI" "toEntity" method

The unit test verifies if the "***toEntity***" method of the "***MapperOpenAiAPI***" class successfully converts data from a model object to a "***ChatCompletionEntity***").

4.6 Technology Review

In this section, we present the technology review for this project. Considering the mutable nature of Software Development, there are several alternatives for frameworks, programming languages, software and architectural patterns, and platforms. One of the requirements considered was cross-

platform compatibility, although this was considered non-prioritary. Nonetheless, designing an application that can be easily adapted to other platforms implies some conditions and limitations on the technologies used. In the following sections, we describe the factors that influenced the choices of programming language and the LLM used in this study.

4.6.1 Programming Language

Considering the creation of an App that may have development across several platforms, while also incorporating VN mechanics with AI integration, we considered four programming languages: Dart, Python, C++ and C#. These were assessed using four criteria: Cross-platform compatibility, Ease of UI development, Performance, and API integration; excluding the use of external Game Engines (e.g., RenPy, Unreal Engine, Unity, Flame Engine).

Regarding the cross-platform compatibility, the languages were assessed in regard to their ability to support development across multiple platforms:

- Dart with Flutter enables straightforward development for several platforms, and compiles to native code;
- Python requires additional frameworks (for VN Games Ren'Py can be used), adding dependency complexity;
- C++ through the use of custom engines enables cross-platform deployment, but requires platform-specific recompilation and adaptation;
- C# through Unity allows cross-platform deployment, but introduces external dependencies;

Regarding cross-platform compatibility, we consider Dart for its ability to simplify the development process for cross-platform.

For the Ease of UI development, we considered the following:

- Dart with Flutter allows for a more straightforward UIs creation than the other options;
- Python and C++ lack native UI tools, depending on frameworks, thus increasing complexity ;
- C# provides UI tools, but has increased complexity.

Dart with Flutter simplifies the process for UI development.

For the Performance, we assessed them considering compilation times and rendering:

- Dart using ahead-of-time (AOT) compilation has fast compilation times and rendering performance while maintaining development simplicity;
- Python is slower in rendering and handling UIs;
- C++ has a higher performance, but has increased complexity;
- C# has a high performance, but has a higher memory usage.

Dart shows a balance between speed and simplicity, making it the choice for applications where efficiency and ease of development are required.

For the API integration, we considered the following:

- Dart provides an easy way to use the HTTP client and the WebSocket support, making API integration seamless;
- Python has extensive libraries for AI, but weaker UI performance;
- C++ can interact with APIs, but has higher complexity;
- C# has HTTP client libraries, but requires more work for AI integration.

Considering only the ease of use for AI, Python would be the better choice. However, considering the AI requirements are simple, Dart is chosen.

Considering these, Dart was chosen as the programming language for the prototype. This choice was made considering: its ease for cross-platform applications; its ease for UI development with Flutter; its balanced performance; and its ease in integrating APIs. For this work, the prototype was deployed and tested only on Android devices.

4.6.2 Large Language Models

Regarding the narrative generation, we considered two options: a local LLM (Ollama/LLama2) and a Remote LLM (OpenAI's GPT models, GPT4o and GPT4o-mini). This choice has a direct influence on the prototype, as the chosen LLM not only impacts narrative generation but also performance, accessibility, cost and privacy. Our decision was based on the following factors:

- Performance:
  - Local LLMs generally have faster response times once loaded, but may suffer with significant initial loading times. Performance depends on the device, which may limit usability.

- Remote LLMs, have faster and consistent response times, but introduce network latency, which may impact real-time narrative generation;

– Cross-platform complexity:

- Local LLMs need manual setup for each platform, with increased difficulty for mobile and web platforms;

- Remote LLMs, accessible through the use of APIs, allow for a simpler setup and only needs minimal alterations for cross-platform. However, they need a constant internet connection;

– Privacy and security:

- Local LLMs offer privacy and security, the data remains on the device;

- Remote LLMs involve sending the data to the remote servers, adding potential risks. However, considering the security and privacy policies of OpenAI and that there is only minimal information from the participants, we consider the risk of personal information breaches is low.

– Cost and maintenance:

- Local LLMs have costs related to hardware requirements, energy consumptions, the cost of maintaining hardware, and the need for software maintenance and updates;

- Remote LLMs remove the need for maintenance and other costs for the hardware and energy consumption, but add costs for usage;

Based on these factors, we chose to use Remote LLMs, considering the need for a simpler cross-platform, and a reduced maintenance and updates burden. Regarding the costs to use the available remote models, the “GPT4o” (prices per 1 Milion tokens, Input: $2.50, cached Input: $1.25, Output: $10.00) has an increased cost compared to the “GPT4o-mini” (prices per 1 Milion tokens, Input: $0.15, cached Input: $0.075, Output: $0.60) model. After some test iterations, we found no significant differences in the generated narratives, thus we chose the “GPT4o-mini” model as the most suitable model for this prototype, with each full game narrative and iteration costing less than or equal to $0.01.

4.7 Requirements

Considering the game concept and the goal of creating a cross-platform playable game, we defined a set of nine functional requirements and ten use cases. These were prioritised according to their relevance for the evaluation of the project outcomes and for answering the research questions.

Table 1 presents the requirements and their implementation status.

Table 1: Requirements Status

Code	Requirement	Implementation Status
FR1	The system must provide a login system that allows players to access the game through their account.	Not Implemented.
FR2	The system must provide a main menu that allows players to start a new game, load a saved game, access settings, logout and/or exit the game.	Partially Implemented – Implemented the main Menu with a New Game option (prioritary for the testing of the outcomes related to the objectives), but the "load a saved game", the "access settings", the "logout and/or exit the game" were not implemented as they are non-prioritary for the evaluation of the outcomes related to the project objectives nor for answering the research questions (see Chapter 1).
FR3	The system must support a save and load system, enabling players to save their progress at any point during the game and load from saved points.	Not Implemented.
FR4	The system must allow players to adjust settings (such as sound volume, text speed, and display settings).	Not Implemented.
FR5	The system must warn players about the sensitive nature of the topics to be presented.	Implemented.
FR6	The system must incorporate branching story paths (determined by the LLM) that enable players to make choices influencing the outcome of the game.	Implemented.
FR7	The system must be able to regenerate a portion of the story if the player requests it.	Not Implemented.
FR8	The system must consider narrative guidelines and rules when creating the narrative or presenting player choices.	Implemented.
FR9	The system must provide a tutorial that explains how the player can interact with the game.	Not Implemented.

As displayed in Table 1, only the requirements directly related to the evaluation of the outcomes of the project (e.g., FR2, FR5, FR6, FR8) were fully or partially implemented. The remaining

requirements were considered non-essential for the evaluation of the outcomes related to the project objectives or for answering the research questions (see Chapter 1) and, thus, are left for future work.

Also, ten use cases were defined based on the requirements. Table 2 presents these use cases and their implementation status.

Table 2: Use Cases Status

Code	**Use Case**	**Implementation Status**
UC1	Create a New Account	Not implemented.
UC2	Login	Not implemented.
UC3	Start New Game	Implemented.
UC4	Save Game Progress	Not implemented.
UC5	Load Game	Not implemented.
UC6	Make Dialogue Choice	Implemented.
UC7	Adjust Settings	Not implemented.
UC8	Access Help or Tutorials	Not implemented.
UC9	Logout	Not implemented.
UC10	Exit Game	Not implemented.

Table 2 shows that only the use cases directly related to the evaluation of the outcomes of the project (UC3, UC6) were implemented. The others were considered non-essential for the evaluation of the outcomes related to the project objectives or for answering the research questions (see Chapter 1) and, thus, are left for future work.

The implementation decisions for both the requirements and use cases were made to focus development on the project's objectives and research questions. This approach allowed for a more focused evaluation of the project's core capabilities while identifying the secondary features for future work.

5 Implementation

This Chapter presents the implementation of the prototype, based on the design considerations presented in Chapter 4. For this project a VN game prototype for Android was developed, where the users interact with the narrative through the "**New Game**" function and progress by selecting choices or moving to the next scene.

This project follows Reso Coder's Clean Architecture (see Section 4.3), ensuring the separation of concerns. State management is managed using BLoC to separate the business logic from the UI and controls the UI state transitions triggered by user events. This application also connects to an LLM for the narrative generation, using TDD to maintain modularity and correctness, with tests written before the implementation and minimal code developed to pass the tests.

The following sections show the Project and Clean Architecture, the use of BLoC for State Management, and the Integration of the LLM for Narrative Generation.

5.1 Project Structure and Clean Architecture

The project structure was designed considering clean architecture, and the TDD approach. At the root, the project is divided into the *lib* and *test* folders, with a sub-folder *features*, where each feature is organised considering the architectural layers. These layers are represented by the *domain*, *data*, and *presentation* folders, as shown by Figure 21.

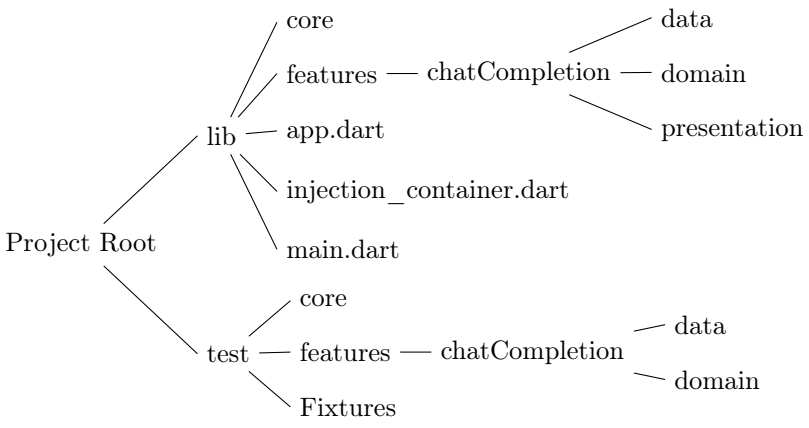


Figure 21: Project Folder Structure

Inside each layer, additional subfolders identify each main component. For example, the domain layer contains the folders: *entities*, *repositories*, and *usecases*. This structure assists in the division of concerns, but also helps in finding the tests created for each class in the *test* folder.

Each of these layers has a purpose:

- Presentation layer: controls the UI, through the presentation logic holders and widgets. BLoC was used, converts user-triggered events to appropriate states that update the UI.
- Domain layer: encompasses the business logic of the application, defining entities, use cases and repository interfaces, making sure the domain logic is independent of the data layer.
- Data layer: controls the access to data and the conversion. This layer comprises the data sources, the models that convert the raw data to the data format used, and repository implementations that control when and how to fetch or save data.

The interaction in these layers follows a clear flow (see Figure 22). When the user triggers an event in the UI, the BLoC converts it to a state, and then either updates the UI or calls a use case. The use case interacts with the repository interface in the Domain layer, which passes it to the data layer. Depending on the situation, the repository will return the data or a "***Failure***", both of which propagate back through the use case and BLoC to update the UI.

Considering the TDD approach, first, writing the test, then implement the minimal code to pass the test, and lastly, refactor while maintaining the tests valid. This reduces the possibility of bugs and ensures that each component tested has only one responsibility.

5.2 State Management Implementation

As identified in Section 4.4, the application uses the BLoC pattern to manage state. The BLoC architecture is composed of three elements: the UI, the BLoC and the data layer (repository and data provider). It has two operating concepts: events that are the user inputs, and states that represent the condition of the application. The BLoC receives events and converts them to the corresponding states.

For this project, we created the following events:

- "***NewGameEvent***", triggered when a player starts a new game;
- "***LoadGameEvent***", triggered when a saved game is loaded;

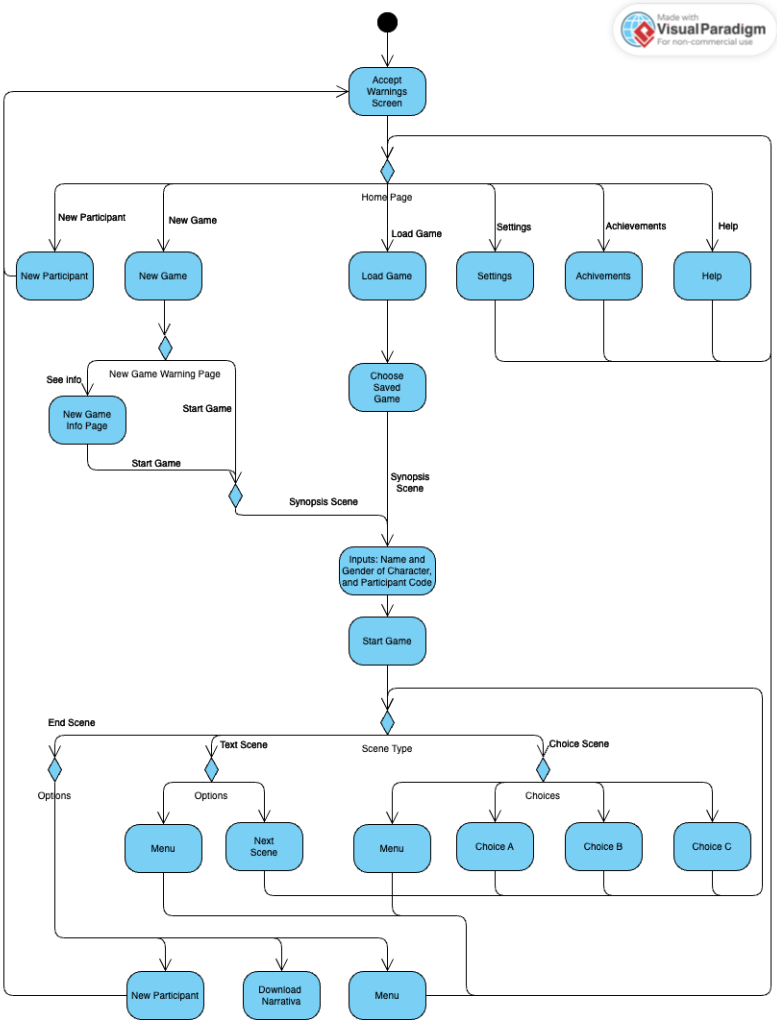


Figure 22: CollegeTales FlowChart

– "***StartSceneEvent***", triggered when the player provides the main character's information and starts the narrative;

– "***NextSceneEvent***", triggered when the player selects the *Next* button in a Text scene;

– "***NextChoiceSceneEvent***", triggered when the player selects one of the choice buttons in a Choice scene.

For the states created to control the UI, we defined them as follows:

– "***Initial***", default state;

– "***LoadingScene***", waiting for the narrative generation, showing a loading widget;

– "***LoadedSynopsisScene***" displays the Synopsis Scene widget;

– "***LoadedTextScene***" displays the Text Scene widget;

– "***LoadedChoiceScene***" displays the Choice Scene widget;

– "***LoadedEndScene***" displays the End Scene widget;

– the last state "***Error***" indicates an error occurred, and shows the Error Screen widget.

The conversions of the events to states follows a pattern. When an event is triggered, the BLoC emits a "***LoadingScene***" state and makes a request to the LLM. Depending on the response, the BLoC emits the appropriate state, e.g., "***LoadedSynopsisScene***" for the "***NewGameEvent***", or one of the other scene states for subsequent events, depending on the scene type received. If the LLM response is an error or an unexpected scene type, a "***Error***" state is emitted, guaranteeing the UI is synchronised with the narrative flow and displays the correct widgets to the player.

5.3 Implementation of Large Language Model Integration for Narrative Generation

Considering the design decisions presented in Section 4.6, the LLM was integrated as a remote data source through the OpenAI API, using the GPT4o-mini model for narrative text generation. Also, the same API was used before the testing to generate images, which were included in the application's assets and displayed in scenes when the LLM selected them, considering a predefined list (for the images, see Appendix N).

The calls to the API were made through a dedicated remote data source class. Requests were made through the HTML Endpoints, with configuration variables extracted from an environment file. The variables included: *model*, *temperature*, *topP*, *frequencyPenalty*, *presencePenalty*, *seed*, *APIURL*, and *APIKey*. Table 3 shows the values used for these variables, excluding the *APIKey*.

The values for *temperature* and *topP* were refined through iterative testing, starting from reference values suggested in the OpenAI Developer Community's "*Cheat Sheet: Mastering Temperature and Top_p in ChatGPT API*"[13]. While the referenced values for *temperature* and *topP* are 0.7 and 0.8 respectively, these were adjusted to try to reduce hallucinations and giving correct mental health information and also maintain narrative consistency, so as to reduce the possibility of high

[13]https://community.openai.com/t/cheat-sheet-mastering-temperature-and-top-p-in-chatgpt-api/172683

Table 3: LLM Parameters Values

Variable	Value	Range	Default
APIURL	https://api.openai.com/v1/chat/completions	—	—
model	"gpt-4o-mini"	—	—
temperature	0.9	0.0 to 2.0	1.0
topP	0.0	—	1.0
frequencyPenalty	0.0	−2.0 to 2.0	0.0
presencePenalty	0.0	−2.0 to 2.0	0.0
seed	52	—	—

variability between narratives as this would be another variable that could affect the outcomes. With a final value of 0.9 for the *temperature* and a value of 0.0 for the *topP*.

The data source verifies every API response by checking the status code. A status code of 200 indicates a successful response of a generated scene, which is then given to a model responsible for converting the API response to the application's data format. A status code different from 200 results in a "*ServerException*", later converted to a failure, and propagated through the BLoC component, making sure that the UI shows the corresponding error message to the player (see Source Code 8).

```
104   Future<ChatCompletionModel> getChatCompletion(
105     List<Map<String, String>> promptRequest, [
106     Map<String, dynamic>? responseFormat,
107   ]) async {

130     if (response.statusCode == 200) {
131       return ChatCompletionModelFactory.getModel(
132         json.decode(utf8.decode(response.bodyBytes)),
133         DataSourceType.remote,
134         CompletionProviders.openAI,
135       );
136     } else {
137       throw ServerException(
138         errorCode: response.statusCode,
139         errorMessage: response.reasonPhrase!,
140         stackTrace: StackTrace.current.toString(),
141         response: response,
142       );
143     }
144   }
145 }
```

Source Code 8: Remote Data Source API Response

Considering that the narrative was generated scene by scene, it was necessary to maintain context without exposing any private user data. Only the character's name, gender, and participant code are transmitted to the LLM. After the initial interaction for the user inputs (character's name, gender, and participant code), subsequent interaction occurred through buttons (e.g., "Next" or scene choices like "Choice A", "Choice B" or "Choice C"), avoiding user inputs. To maintain the context and provide it to the LLM, each new prompt is appended to a list containing all previous prompts and generated scenes, ensuring narrative progression and consistency.

6 Evaluation

6.1 Study Design

We conducted an empirical mixed-methods research study with the objective to evaluate whether the VN prototype impacted MHL in CSs. The study was conducted within the academic community of the first author's institution. Participants were selected through convenience sampling and had to meet the following criteria: (1) they needed to be CSs aged 18 or older, (2) they needed to have English reading proficiency, and (3) they needed to be capable of interacting with a mobile game. Each participant was assigned a unique identification code (e.g., P11, P02). The primary outcomes that were measured included DL and MiscD, alongside qualitative insights from interviews. Additionally, the study was conducted without professional supervision considering the lack of a psychological or mental health team, limiting the control.

In terms of the participants, this study used a sample of 28 CSs, and consisted of non-native English speakers. The sample size was calculated using the G*Power software [66], using a one-tail a priori calculation. Considering the comparison of pre- and post-intervention data and a Cohen's d effect of 0.65, the total number of participants calculated was 28. The participants' ages ranged between 22 and 41 years (mean M=27.29, Standard deviation SD=5.01), of these, 11 are male, 16 female, and 1 other. Table 4 shows the general data of the participants, while Table 6 shows the gender distribution of the sample.

6.2 Protocol

Before the start of the experiment with the college students, and considering the sensitive theme in the study, the study was sent to the Data Protection Officer from the University of Madeira and subsequently to the Ethics Committee from the same institution for their opinion, receiving approval from both entities (see Appendix A for the Ethics Committee). Regarding the Pilot Study with the MH professionals, considering its exploratory method to obtain feedback from the professionals, it wasn't sent to the Data Protection Officer nor to the Ethics Committee. After the approval, we started the experiment, which was conducted following these steps:

1. Participants were introduced to the study and prototype, gave informed consent (see Appendix E), and received two documents: Mental Health support contacts and a link for a follow-up questionnaire.

Table 4: Participants Sociodemographic Data

Identifier Code	Age	Gender	Academic Level
P01	23	Female	Bachelor Degree
P02	28	Male	Masters Degree
P03	29	Male	Masters Degree
P04	40	Male	Post-Graduate
P05	27	Male	Professional Higher Tecnhical Courses
P06	25	Female	High School
P07	25	Female	Bachelor Degree
P08	24	Female	Masters Degree
P09	30	Female	Masters Degree
P10	25	Female	Bachelor Degree
P11	28	Female	Masters Degree
P12	26	Female	Bachelor Degree
P13	24	Male	Masters Degree
P14	30	Female	Masters Degree
⋮	⋮	⋮	⋮
P15	39	Female	Masters Degree
P16	23	Female	Bachelor Degree
P17	41	Female	High School
P18	28	Female	Bachelor Degree
P19	25	Male	Bachelor Degree
P20	27	Female	Bachelor Degree
P21	25	Male	Bachelor Degree
P22	25	Male	Bachelor Degree
P23	22	Male	Bachelor Degree
P24	22	Male	Bachelor Degree
P25	24	Female	Masters Degree
P26	29	none	Masters Degree
P27	25	Male	Bachelor Degree
P28	25	Female	Masters Degree

Table 5: Sample Academic Level Statistics

Academic Level	Frequency	Percentage
High School	2	7.14
Professional Higher Tecnhical Courses	1	3.57
Bachelor Degree	13	46.43
Masters Degree	11	39.29
Post-Graduate	1	3.57

Table 6: Sample Gender Statistics

Gender	Frequency	Percentage
Female	16	57.14
Male	11	39.29
Other	1	3.57

Table 7: Descriptive statistics for Age

	M	SD	Min	Q1	Median	Q3	Max
Age	27.29	5.01	22	24.75	25	28.25	41

2. Each participant was assigned a unique participant identification code to anonymise their data (e.g., 'PXX').

3. Participants then were asked to complete pre-test questionnaires (sociodemographics questionnaire and Depression Literacy and Misconceptions Scale (DepSter Scale) questionnaire), in person via computer, or remotely via screen-sharing.

4. Participants were then asked to test the prototype, either in person via computer, or remotely via screen-sharing. Participants also consented to the collection of their anonymised generated narratives.

5. After the testing of the prototype, the participants were asked to answer a post-test questionnaire including the Narrative Transportation Scale (NT Scale) and the DepSter Scale, either in person via computer, or remotely via screen-sharing.

6. Participants, after consenting to being recorded, participated in a semi-structured interview, which was recorded using a voice-distortion App to ensure the participants' anonymity.

7. At the one-week follow-up, consenting participants were emailed a follow-up questionnaire link (also provided at the start through a supporting document).

6.3 Measures

A mixed methods approach was employed in this research. Considering the research questions presented in Chapter 4, we applied a qualitative method to evaluate RQ1.

Qualitative data were obtained from the semi-structured interviews performed at the post-test, which were designed to explore: (1) The participants' previous experience with similar applications, (2) the narrative impact on the participants' knowledge (literacy and misconceptions), stigma, and emotional response, and (3) the participants' perceptions of using LLMs for MHL interventions.

Regarding RQ2, we employed a quantitative method to assess whether the prototype affects any of the outcomes: DL, MiscD, and Narrative Transportation (NT). Through this analysis, we evaluate, in three timepoints, whether the prototype significantly impacts the participants' outcomes and whether there is any significant retention of knowledge for the DL and/or MiscD scores, in the short term (1 week) (third timepoint). In this approach, we used two scales and a socio-demographics questionnaire (Age, Gender and Completed Academic Level), applied before the test of the prototype (first timepoint). The two scales applied, are the NT Scale, in english, by Green and Brock [67] to evaluate the NT, used after the test of the prototype (second timepoint) and

the DepSter Scale, in it's english version, by Kulwicka and Gasiorowska [68] to evaluate the DL and the MiscD, applied at the three timepoints for data collection, at pre-test, post-test, and one-week follow-up. To evaluate the effect of the prototype on literacy and misconceptions about depression, and the existence or not of retention of knowledge.

Regarding the NT variable, we used the NT Scale on its original version, with 11 general items, and 1 or more imagery items, considering the number of main characters in the story, including the following example for an item "*The narrative affected me emotionally*". The scale results were obtained considering a 7-point Likert Scale (coding the scores from 1 = "*not at all*" to 7 = "*very much*").

We adapted the scale to the narrative each participant received, thus, the scale had 11 general items and only one imagery item, as each narrative has only one main character. This adaptation was made considering that the original scale states that item 12 can be duplicated for each main character in a story. However, this adaptation treats the scale as a unidimensional construct, rather than separating it into the three dimensions originally reported (cognitive, affective and imagery). The original scale also states that these three dimensions (cognitive, affective and imagery) do not predict relevant outcomes differently, and thus the authors reported the results as only one construct ("full scale") [67]. Therefore, this approach may be justified, and for this unidimensional construct, the possible score range was from 12 to 84.

Considering this, we calculated the Cronbach's Alpha (internal consistency) for the sample in this research, reporting a value of $\alpha = 0.76$, with a confidence interval of 95% $[0.604, 0.873]$. The Cronbach's Alpha reported by Green and Brock [67] for the original scale with 11 general items, and 4 imagery items specific to the narrative used in the original validation was $\alpha = 0.76$. While Cronbach's alpha is reported, McDonald's Omega can be used as an alternative. We calculated the McDonald's Omega for the sample, reporting a value of $\omega = 0.86$.

As for the item-dimension correlations, considering the unidimensional construct, we report the correlations in Table 8.

For the variables DL and MiscD, these were evaluated as two dimensions on the DepSter Scale, through the use of its English version, validated by Kulwicka and Gasiorowska [68], with the full scale containing fourteen items, the scores of each item were obtained considering a 5-point Likert Scale (coding the scores from 1 = "*Strongly disagree*" to 5 = "*Strongly agree*").

Table 8: Item-Dimension Correlation NT Scale

Item	**Correlation item-total Postest**
"While I was reading the narrative, I could easily picture the events in it taking place."	0.61
"While I was reading the narrative, activity going on in the room around me was on my mind."	0.17
"I could picture myself in the scene of the events described in the narrative."	0.50
"I was mentally involved in the narrative while reading it."	0.70
"After the narrative ended, I found it easy to put it out of my mind."	0.20
"I wanted to learn how the narrative ended."	0.75
"The narrative affected me emotionally."	0.47
"I found myself thinking of ways the narrative could have turned out differently."	0.33
"I found my mind wandering while reading the narrative."	0.22
"The events in the narrative are relevant to my everyday life."	0.46
"The events in the narrative have changed my life."	0.62
"I had a vivid mental image of the Main Character [character name]."	0.67

Considering the first subscale, which assesses the DL dimension, comprised of six items, respectively items 1, 3, 5, 7, 9, and 11, an example is the following item "*Depression can affect anyone*". The scores for this subscale, evaluating the DL dimension, were obtained by averaging the scores of the six items. The possible score range for this subscale was from 1 to 5.

For the MiscD dimension, the subscale contains the remaining eight items, respectively items 2, 4, 6, 8, 10, 12, 13, and 14. An example is the following item "*Depression is just a fad*". The scores for this subscale, evaluating the MiscD dimension, were obtained by averaging the scores of the eight items. The possible score range for this subscale was from 1 to 5.

Considering this, we calculated the Cronbach's Alpha (internal consistency) for the sample in this research. We present in Table 9 the Cronbach's Alpha for each subscale in each of the three timepoints, accompanied by their confidence intervals, and the Cronbach's Alpha's values reported by Kulwicka and Gasiorowska [68], for the original validations of the English scale, in both a US Sample (Study 4), and a UK Sample (Study 5). Considering the sample used was obtained through convenience sampling, the sample consisted of non-native English speakers, possibly influencing the internal consistency values obtained in this study, as their native language and culture can influence how they understand and answer to the items.

Table 9: Cronbach Alphas for the measured Constructs at the three timepoints

Construct	Timepoint	Cronbach's Alpha	95% CI (Lower)	95% CI (Upper)	Cronbach's Alpha DepSter (Study 4)	Cronbach's Alpha DepSter (Study 5)
DL	Pre-test	0.346	-0.118	0.659	0.59	0.49
	Post-test	0.523	0.185	0.752		
	Follow-up	0.713	0.510	0.851		
MiscD	Pre-test	0.405	0.002	0.687	0.93	0.85
	Post-test	0.652	0.417	0.817		
	Follow-up	0.650	0.413	0.816		

While Cronbach's alpha is reported, McDonald's Omega can be used as an alternative. We calculated the McDonald's Omega for the sample, reporting the values for: DL Pre-test: $\omega = 0.73$; DL Post-test: $\omega = 0.81$; DL Follow-up: $\omega = 0.83$; MiscD Pre-test: $\omega = 0.67$; MiscD Post-test: $\omega = 0.83$; and MiscD Follow-up: $\omega = 0.82$.

As for the item-dimension correlations, considering each dimension, we present the correlations in Table 10.

6.4 Analysis

The data collected in this study were the qualitative data from the semi-structured interview and the generated narratives, and the quantitative data from the NT Scale and the DepSter Scale. The qualitative data from the semi-structured interviews were analysed through a Thematic Analysis conducted through Taguette[14], based on the approach described by Braun and Clarke [63]. One researcher coded the transcripts inductively, subsequently extracting, reviewing and refining the results by the first and last authors. The generated narratives were analysed to verify the consistency between the generated narratives, showing that no unanticipated situation arose that contrasted with the original prompt or the previous testing. The quantitative data were analysed statistically, using a repeated measures method (dependent on the time point of application of the sample, before the prototype or pre-test, after the prototype or post-test, and one week after or one-week follow-up) with a dependent sample. Due to non-normal data distributions in DL at the one-week follow-up timepoint (see Table 12), we chose to use non-parametric tests for all variables. Considering the small sample size, we applied the Friedman test, using the F-test adjustment to address the chi-square limitations in small samples, particularly near the tails [69].

[14] `https://www.taguette.org/`

Table 10: Item-Dimension Correlation DepSter Scale

Item	**Correlation Pre-Test**	**Correlation Post-Test**	**Correlation 1-Week Follow-up**	**Original Scale Correlation**
Depression Literacy				
"Depression is an illness"	0.43	0.47	0.72	**0.32**
"Depression can affect anyone"	0.50	0.49	0.18	**0.33**
"Depression makes people lose interest even in the things they used to enjoy doing"	0.53	0.50	0.80	**0.43**
"Depression makes people lack the strength to do any-thing"	0.53	0.54	0.75	**0.42**
"People with depression often think about suicide"	0.45	0.54	0.80	**0.36**
"Depression is associated with great suffering"	0.59	0.75	0.53	**0.41**
Misconceptions about Depression				
"Depression is just a fad"	0.54	0.79	0.49	**0.50**
"Depression is just a temporary mood deterioration"	0.28	0.53	0.26	**0.34**
"Depression affects only the people who are weak and cannot cope with their life"	0.48	0.55	0.64	**0.51**
"To overcome depression, all you need is willpower"	0.53	0.57	0.61	**0.68**
"To overcome depression, all you need is to get yourself together"	0.63	0.75	0.82	**0.70**
"Antidepressant medication start to work right after the intake"	-0.02	0	0.29	**0.27**
"Depression is just a self-pitty"	0.55	0.64	0.78	**0.53**
"People with depression are mentally weak"	0.57	0.61	0.61	**0.47**

7 Results

In this Chapter, we present the study's findings, divided into two: quantitative and qualitative analyses.

7.1 Quantitative Data

This Section will present the Descriptive Statistics, Normality Tests, Inferential Statistics, and Exploratory Analyses/Correlations.

Descriptive statistics for the DL, MiscD, and NT scores collected at the different timepoints are presented in Table 11.

Scores remained stable on all timepoints for DL, while MiscD decreased slightly after the intervention. NT scores suggest a moderately high engagement with the prototype.

Table 11: Descriptive statistics for Depression Literacy, Misconceptions of Depression, and Narrative Transportation scores collected at different timepoints

Score Type	**Timepoint**	**M**	**SD**	**Min**	**Q1**	**Median**	**Q3**	**Max**
Depression Literacy[1]	Pre-test	4.16	0.40	3.50	3.96	4.17	4.38	5.00
	Post-test	4.18	0.42	3.50	3.83	4.17	4.38	5.00
	Follow-up	4.14	0.58	2.50	3.67	4.08	4.67	4.83
Misconceptions of Depression[2]	Pre-test	1.81	0.40	1.25	1.50	1.75	2.13	2.63
	Post-test	1.63	0.40	1.00	1.38	1.63	1.78	2.38
	Follow-up	1.66	0.41	1.00	1.34	1.69	1.91	2.63
Narrative Transportation[3]	Post-test	56.68	7.91	38.00	51.75	57.50	62.25	70.00

[1] Lower scores imply lower Literacy, while higher scores imply more.
[2] Lower scores imply lower Misconceptions, while higher scores imply more.
[3] Possible score range from 12 to 84.

Normality of the data was assessed using the Shapiro-Wilk test, as shown in Table 12.

Almost all variables followed a normal distribution, except for the Follow-up DL scores ($p = 0.020$). Given this violation of normality and the small sample size, non-parametric tests were employed for all variables (including MiscD that is normal) in the inferential analysis. Nevertheless, even when applying a parametric method, specifically repeated measures ANOVA, the variable MiscD would still be statistically significant.

A Friedman test showed no statistically significant differences in DL scores across time points ($F = 0.078$, $p = 0.920$). However, there was a statistically significant difference in MiscD, $F =$

Table 12: Shapiro-Wilk test results for normality of Depression Literacy, Misconceptions of Depression, and Narrative Transportation scores across timepoints

Score Type	Timepoint	W	p-value	Normal
Depression Literacy	Pre-test	0.961	0.376	Yes
	Post-test	0.956	0.272	Yes
	Follow-up	0.911	0.020	No
Misconceptions of Depression	Pre-test	0.941	0.115	Yes
	Post-test	0.955	0.263	Yes
	Follow-up	0.969	0.566	Yes
Narrative Transportation	Post-test	0.976	0.755	Yes

6.155, $p = 0.004$. Post hoc analysis with Wilcoxon signed-rank test was conducted with Holm-Bonferroni correction for multiple comparisons. Median (IQR) MiscD levels were 1.750 (1.500 *to* 2.125), 1.625 (1.375 *to* 1.781) and 1.688 (1.344 *to* 1.906), respectively (see Table 11). There were no significant differences between the pre-test and 1-week follow-up ($W = 80.5$, $p_{unc} = 0.047$, $p_{corr} = 0.093$) or between the post-test and 1-week follow-up ($W = 103.5$, $p_{unc} = 0.684$, $p_{corr} = 0.684$). However, there was a statistically significant reduction in MiscD in the pre-test to post-test ($W = 37.0$, $p_{unc} = 0.002$ $p_{corr} = 0.006$).

Table 13: DepSter Scale Constructs Ranks

Construct	Timepoint	Sum of Ranks	Mean Ranks
Depression Literacy	Pre-Test	55.5	1.98
	Post-Test	55.0	1.96
	Follow-up	57.5	2.05
Misconceptions of Depression	Pre-Test	68.5	2.45
	Post-Test	47.0	1.68
	Follow-up	52.5	1.88

Considering MiscD has a normal distribution, a parametric method could be used with it, and it would still be statistically significant. By using ANOVA, there was a statistically significant difference, $F = 5.199$, $p = 0.009$. Post hoc pairwise comparisons with a Holm-Bonferroni correction applied, resulting in a significance level set at $p < 0.017$) There were no significant differences between the pre-test and 1-week follow-up ($T = 2.135$, $p = 0.042$) or between the post-test and 1-week follow-up ($T = -0.548$, $p = 0.588$). However, there was a statistically significant reduction in MiscD in the pre-test to post-test ($T = 3.402$, $p = 0.002$).

Figure 23 presents Spearman's correlations for the demographic variables and measured outcomes.

The exploratory correlation analysis revealed several patterns. DL scores at different timepoints were strongly positively correlated (e.g., Pre-test and Post-test: $r = 0.8052$), while MiscD were negatively correlated with DL at all timepoints (e.g., Post-test MiscD and DL: $r = -0.5621$). NT showed a weak negative correlation with Post-test MiscD ($r = -0.3403$), suggesting greater immersion may be associated with fewer misconceptions. Gender showed a moderate negative correlation with 1-week follow-up MiscD scores ($r = -0.4274$), which may justify further exploration.

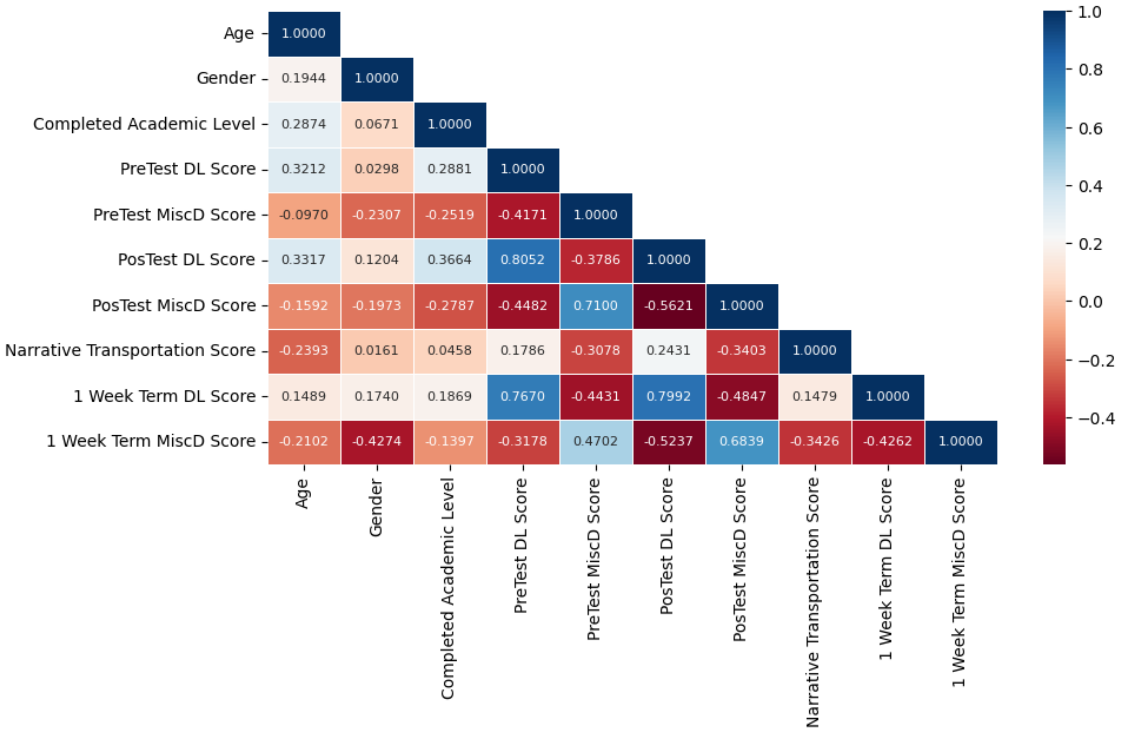


Figure 23: Correlations between Scores, Age, Gender, and Completed Academic Level

7.2 Qualitative Data

Through the thematic analysis of interview transcripts, we identified four high-level themes: "Digital Tools for Mental Health Education and Their Potential", "Authenticity and Emotional Engagement with Narrative", "Social Attitudes and Stigma Toward Mental Health", and "Trust and Ethical Considerations in AI-Generated Narratives".

Participants indicated that they didn't have experience with digital tools to promote MHL. However, many would be interested in trying if they felt it was relevant or personally needed. (P07

stated "*If I knew any, like, more relevant, i think so*"), or if they needed it (P06 stated "*If I need yes[...]*").

Several participants also reported how the VNs narrative had an impact on their MHL and knowledge, while also being beneficial in real-world settings. P14 emphasised its potential for use by professionals "*Yes, they could [be used by Mental Health professionals], and that is very interesting. Why? Because it is completely different for us to give this narrative to someone who is not from the mental health field and then not know what answer to give, right? Than to give it to a professional in this field who is there and can provide support.*" P02 reflected on its personal impact, describing it as a way to practice emotional decision-making and testing the consequences of their actions: "*Because it is a way for you to test the cause and effect of your actions in a virtual environment. [...] And the fact that you test it in a virtual environment and see that it really worked can also lead you, in real life, to make the same decisions and see that it's not so bad.*").

Participants noted that the generated narratives were authentic, relatable, and emotionally engaging. P12 stated, "*It seemed authentic to me. It even seemed like a situation I went through with a friend, but that wasn't the way I went with her, and maybe it could have been.*" These narratives also evoked emotional responses and reflection. P11 stated "*[...] I could understand that situation[...] I started to get really happy. [...] even though I think like, "but that's stupid, because I'm not doing anything, [...] I think that unconsciously I know that this is a story that could be happening to someone. [...]*" Others, like P01, emphasised the narratives also evoked empathy: *"Yeah, because it made me see once again the first person perspective of having those feelings, so it made me feel empathy towards them."*

Several participants noted that the narrative showed some stressors and social pressures unique to CSs. P14 stated, *"Yes, it presents some pressures that other classes do not have, as is the case when he talks about exams [...] it is clear that they are university students given what is presented [...]"*

Participants also highlighted that these narratives could help reduce stigma around MH. P11 stated common misconceptions: "*[...] Because there is still that idea that 'you are putting on a show', 'you are making a movie', 'it's all in your head', 'you are exaggerating' [...]. And I think it is the opposite. I think that depression affects everyone, yes, it mainly affects those who others see as being too strong.*". However, participants also mention that personal openness plays a significant

role. As P11 stated, "*Yes, if they wanted to learn about depression, but didn't have stigma about depression. [...] Because this helps those who really want to learn.*". Also, P26 stated scepticism about the impact: "*But I don't know like if people I know that have this kind of wrong ideas about mental health or about depression, could actually change their minds. Because I think even when they have somebody close and they see them suffering, this doesn't really mean they change much. Because I think it's more like on your own attitude towards change that you change than actually like through the narrative itself. So, yeah, I'm unaware.*".

While participants generally acknowledged that the narratives were coherent and engaging, concerns were raised about the trustworthiness and ethical implications of AI-generated content. P08 questioned whether AI could truly convey emotional nuance: "*[Artificial Intelligence] It is capable of portraying these moments like this, but I don't know if artificial intelligence has the sensitivity that we also have. [...] To talk about some topics, like depression. It has no idea about these things, like it has the literature, but it doesn't have the sensitivity.*".

P26 also stated concerns about emotional authenticity and depth: "*[...] I'm not sure, if I feel as comfortable, like talking to a robot about human psychology, even if like, of course this is based in a great corpus of information, like a great dataset of mental health care and etc, etc. I think I'm more oriented towards crafted narratives, and I believe that crafted narratives can have a wider potential of actually, kind of creating a more complex way of understanding, and maybe more engaging ways. [...] It was very basic, kind of like there is these kind of type subjects, and there is this very kind of type situation, and then you can use this type situation to go to the sub subjects, so it was like very straightforward, I felt.*". These underscore a need for greater transparency and human oversight in systems addressing sensitive psychological issues.

Table 14: Code Groups Experiment

Code Group	**Initial Codes**
CG1 - Use of Digital Tools for Mental Health Literacy	C1 – Never used digital tools or applications for mental health literacy; C2 – Used digital tools or applications for mental health literacy; C3 – Good experience with past used digital tools or applications for mental health literacy; C4 – Interested in trying digital tools or applications for mental health literacy;

Continued on next page

Table 14 — Code Groups Experiment (continued)

Code Group	**Initial Codes**
	C5 – Not Interested in trying digital tools or applications for mental health literacy;
CG2 - Narrative Authentic and Relatable	C6 – Narrative was authentic and relatable; C7 – Narrative was not authentic but was relatable; C8 – Wrong portrayal of recuperation in the narrative; C9 – Narrative was authentic but not relatable; C10 – Narrative was not authentic nor relatable;
CG3 - Impact on the Mental Health Literacy and Knowledge	C11 – Good representation of the reality of Depression on the narrative; C12 – Oversimplification of the reality of Depression on the narrative; C13 – Didn't learn anything new about Depression literacy; C14 – Learned about a first person perspective of who has Depression; C15 – Learned something new about Depression literacy; C19 – Narrative gave common knowledge; C20 – Incorrect Knowledge extraction from the narrative;
CG4 - University Students and Their Unique Situations and Stressors	C16 – Shows University Students unique situations and stressors; C26 – Doesn't show many situations and stressors from university students, it's more general;
CG5 - Use and Applications on Real World	C17 – The narrative was clear and easy to understand; C24 – Several possibilities of narrative should be tried as each can give different knowledge; C31 – The information given by the narrative can be useful in real Life situations; C32 – Would recommend the narrative; C33 – The narrative can be used by Mental Health Therapists; C34 – There are several possible branches on the narrative; C35 – Would recommend the narrative for younger people, but not for older than high school students; C39 – Immersed in the narrative;
CG6 - Stigma and "Openness" to learn	C18 – The narrative can reduce the stigma; C21 – Narrative increases stigma;

Continued on next page

Table 14 — Code Groups Experiment (continued)

Code Group	**Initial Codes**
	C22 – Potential for correction of the narrative to reduce stigma; C23 – People need to be "open" so the narrative can have an effect; C25 – Mental Health Stigma opposed to Physical Health Stigma; C27 – Unsure if the narrative can reduce stigma;
CG7 - Effect on Emotions, Reflections and Empathy	C28 – The narrative evoked emotions and reflections; C29 – The narrative provided a sense of empathy; C30 – The narrative evoked emotions but not reflections; C36 – The narrative did not evoke emotions nor reflections; C37 – The narrative did not evoke emotions but evoked reflections;
CG8 - Limitations and Concerns of AI-Generated Narratives	C38 – Lack of Human Sensitivity on the Artificial Intelligence;

Table 15: Themes and Code Groups Experiment

Theme	**Code Group**
Digital Tools for Mental Health Education and Their Potential	CG1 - Use of Digital Tools for Mental Health Literacy CG3 - Impact on the Mental Health Literacy and Knowledge CG5 - Use and Applications on Real World
Authenticity and Emotional Engagement with Narrative	CG2 - Narrative Authentic and Relatable CG7 - Effect on Emotions, Reflections and Empathy
Social Attitudes and Stigma Toward Mental Health	CG4 - University Students and Their Unique Situations and Stressors CG6 - Stigma and "Openness" to Learn
Trust and Ethical Considerations in AI-Generated Narratives	CG8 - Limitations and Concerns of AI-Generated Narratives

8 Discussion

This work explored the impact of a digital narrative-based intervention on DL, MiscD, and NT on CSs. The evaluation of the VN prototype revealed significant effects on the participants' MiscD.

Statistical analysis using non-parametric tests showed that while DL scores remained high from the pre-test to the follow-up, indicating a high baseline knowledge and suggesting a potential ceiling effect, the intervention significantly reduced MiscD immediately post-intervention, with these improvements generally maintained at follow-up. Correlation patterns supported this result: higher literacy scores were negatively associated with misconceptions, reinforcing the relationship between accurate MH knowledge and reduced stigma.

NT, measured only at post-test, showed moderate to high engagement levels ($M = 56.68$, $SD = 7.91$), and was moderately negatively correlated with MiscD scores, indicating that the digital narrative was successful in engaging and immersing users, which is a key factor in the effectiveness of narrative interventions [67], while also suggesting that greater narrative engagement may help reduce misconceptions. NT also showed a small positive correlation with DL scores at post-test and follow-up, indicating that participants who felt more immersed in the narrative tended to retain higher levels of domain knowledge. This aligns with broader evidence that interactive and immersive storytelling can support behaviour change and learning, especially when grounded in relatable scenarios [52, 67, 70].

The qualitative results support this view. Participants described the experience as emotionally engaging, reflective, and realistic. Several reported that the narrative prompted them to consider the real-world consequences of decisions, empathise with characters, and reflect on their attitudes toward MH. This emotional connection and understanding are critical factors in effective health communication and are often difficult to achieve in traditional, didactic educational approaches [67, 71].

One of the main findings of this study was the participants' ambivalence about the ability of AI-generated narratives to fully capture the emotional depth of human experience, particularly regarding sensitive topics like depression. This concern is valid, especially in sensitive domains like MH.

Answering the research questions, the findings imply that RQ1 – "*How can we leverage large language models to generate narratives in game-based mental health interventions?*", can be answered that the LLMs can be used to generate immersive and interactive, branching narratives while taking into consideration previous scenes and the player choices, and can create engaging narratives even if with limitations regarding simulating the emotional depth that is inherent to humans . As for RQ2 – "*How do game-based mental health interventions affect mental health literacy of young adults?*", the findings suggest that MH interventions can significantly reduce MiscD even when DL is at a high level at baseline.

This work shows that LLMs can be used as creative tools, allowing professionals without programming or storytelling expertise (e.g. MH professionals, educators) to create immersive gamified narrative interventions for MHL. They reduce the technical knowledge needed to create branching narratives. However, while they can be used to create emotionally immersive scenes, there are limitations regarding the language and the emotional depth inherent to humans. Findings showed that digital interventions, through interactive narratives, can reduce MiscD, even if the knowledge levels of DL are high at baseline.

The high baseline of DL was not expected, and limited the potential to observe knowledge gains, creating a ceiling effect. However, the reduction of MiscD suggests that misconceptions may be more malleable than knowledge in this population.

These findings present some implications for future research: (1) the potential of designing LLM generated narratives with human co-design to increase the emotional depth and the control of the given information, answering the concerns about the authenticity; (2) future research should verify the participants' baseline knowledge, and test with a population with lower baseline knowledge; (3) future work could explore the participants Big Five personality traits and analyse if results grouped by personality have similar or different patterns on the outcomes and interaction behaviours; and (4) lastly, future research should explore whether the effects from the MH intervention are maintained in the future and if they affect the participants' behaviour related to MHL, demonstrating the whether these interventions can affect the "action part" of MHL defined by Jorm [43].

Interactive narratives offer immersion, agency, and emotional engagement through the choices (or interactions) presented to players, providing an opportunity for reflection on decision-making

and consequences, and allowing players to simulate complex real-life situations in a controlled environment.

9 Conclusion

Future research should explore ways to sustain reductions in misconceptions, investigate the impact of narrative complexity and authenticity, and examine the integration of human and AI-authored content. Investigations with larger and more diverse samples should be conducted to assess the long-term effectiveness and broader uses of digital narrative interventions in MH education.

Several limitations should be noted. First, the sample size was relatively small, which may restrict the generalizability of the findings. Moreover, the study was conducted in a country where English is not the primary language, and the participants were non-native English speakers interacting with an English prototype, which may have influenced their understanding and responses, which is also one of the possible reasons for the low cronbach alphas reported, for the DepSter Scale, as its an English scale. This study, was also conducted without professional supervision and support by a psychological or mental health team, limiting psychological control, validation, and clinical validity. Additionally, the qualitative sample may not fully represent the diversity of user experiences, as some of the open-ended questions left participants unsure of how to respond. The high DL scores at baseline indicate a potential ceiling effect. NT was only assessed at the post-test, and the absence of a delayed or replicated intervention may have restricted long-term effects. Other limitations included the inability of AI to represent human emotions, the capacity to earn user trust, especially when they are aware that the content is generated by AI, and difficulties in maintaining improvements on the outcomes measured over time.

Overall, while this work has limitations, our findings demonstrate that LLM generated narratives can enhance MHL in young adults. Even with the limitations, LLMs can be powerful tools for creating personalised, engaging, and scalable MH interventions. When applied safely, they can help non-experts create emotionally immersive educational tools, giving way for more inclusive, accessible, and impactful MHL initiatives.

References

[1] World Health Organization, *Basic documents*. World Health Organization, 2020.

[2] ——, *World mental health report: Transforming mental health for all*, 2022. [Online]. Available: https://www.who.int/publications-detail-redirect/9789240049338

[3] A. Wasim, J. Truong, S. Bakshi, and U. Majid, "A systematic review of fear, stigma, and mental health outcomes of pandemics," *Journal of Mental Health*, vol. 32, no. 5, pp. 920–934, Sep. 2023. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/09638237.2022.2091754

[4] S. SeyedAlinaghi, A. M. Afsahi, R. Shahidi, A. Afzalian, P. Mirzapour, M. Eslami, S. Ahmadi, P. Matini, S. Yarmohammadi, S. Saeed Tamehri Zadeh, P. Asili, P. Paranjkhoo, M. Ramezani, S. Nooralioghli Parikhani, F. Sanaati, I. Amiri Fard, E. Emamgholizade Baboli, S. Mansouri, A. Pashaei, E. Mehraeen, and D. Hackett, "Social stigma during COVID-19: A systematic review," *SAGE Open Medicine*, vol. 11, p. 20503121231208273, Jan. 2023. [Online]. Available: http://journals.sagepub.com/doi/10.1177/20503121231208273

[5] D. F. Santomauro, A. M. Mantilla Herrera, J. Shadid, P. Zheng, C. Ashbaugh, D. M. Pigott, C. Abbafati, C. Adolph, J. O. Amlag, A. Y. Aravkin, B. L. Bang-Jensen, G. J. Bertolacci, S. S. Bloom, R. Castellano, E. Castro, S. Chakrabarti, J. Chattopadhyay, R. M. Cogen, J. K. Collins, X. Dai, W. J. Dangel, C. Dapper, A. Deen, M. Erickson, S. B. Ewald, A. D. Flaxman, J. J. Frostad, N. Fullman, J. R. Giles, A. Z. Giref, G. Guo, J. He, M. Helak, E. N. Hulland, B. Idrisov, A. Lindstrom, E. Linebarger, P. A. Lotufo, R. Lozano, B. Magistro, D. C. Malta, J. C. Månsson, F. Marinho, A. H. Mokdad, L. Monasta, P. Naik, S. Nomura, J. K. O'Halloran, S. M. Ostroff, M. Pasovic, L. Penberthy, R. C. Reiner Jr, G. Reinke, A. L. P. Ribeiro, A. Sholokhov, R. J. D. Sorensen, E. Varavikova, A. T. Vo, R. Walcott, S. Watson, C. S. Wiysonge, B. Zigler, S. I. Hay, T. Vos, C. J. L. Murray, H. A. Whiteford, and A. J. Ferrari, "Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic," *The Lancet*, vol. 398, no. 10312, pp. 1700–1712, Nov. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0140673621021437

[6] N. B. Anderson, K. C. Nordal, D. W. Ballard, L. F. Bufka, J. Diaz-Granados, G. P. Keita, C. V. Wright, L. Bossolo, and S. Bethune, "Stress in America: The Impact of Discrimination," 2016.

[7] J. Hunt and D. Eisenberg, "Mental Health Problems and Help-Seeking Behavior Among College Students," *Journal of Adolescent Health*, vol. 46, no. 1, pp. 3–10, Jan. 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1054139X09003401

[8] P. Ratanasiripong, K. Sverduk, D. Hayashino, and J. Prince, "Setting up the next generation biofeedback program for stress and anxiety management for college students: A simple and cost-effective approach," *College Student Journal*, vol. 44, pp. 97–100, Jan. 2010.

[9] J. H. Lee, S. Cho, S. Lee, W. J. Eum, H. Jang, S. Suh, and S. M. Lee, "Initial Validation of the Planned Happenstance Career Inventory–English Version," *The Career Development Quarterly*, vol. 65, no. 4, pp. 366–378, Dec. 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/cdq.12114

[10] R. V. Kadison and T. F. DiGeronimo, *College of the overwhelmed: the campus mental health crisis and what to do about it.* San Francisco, Calif: Jossey-Bass, a Wiley imprint, 2004.

[11] N. J. Reavley, T. V. McCann, and A. F. Jorm, "Mental health literacy in higher education students," *Early Intervention in Psychiatry*, vol. 6, no. 1, pp. 45–52, Feb. 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.1751-7893.2011.00314.x

[12] N. Reavley and A. F. Jorm, "Prevention and early intervention to improve mental health in higher education students: a review," *Early Intervention in Psychiatry*, vol. 4, no. 2, pp. 132–142, May 2010. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.1751-7893.2010.00167.x

[13] L. T. Lam, "Mental health literacy and mental health status in adolescents: a population-based survey," *Child and Adolescent Psychiatry and Mental Health*, vol. 8, no. 1, p. 26, 2014. [Online]. Available: http://capmh.biomedcentral.com/articles/10.1186/1753-2000-8-26

[14] L. M. Loureiro, A. F. Jorm, A. C. Mendes, J. C. Santos, R. O. Ferreira, and A. T. Pedreiro, "Mental health literacy about depression: a survey of portuguese

youth," *BMC Psychiatry*, vol. 13, no. 1, p. 129, Dec. 2013. [Online]. Available: http://bmcpsychiatry.biomedcentral.com/articles/10.1186/1471-244X-13-129

[15] L. Townsend, R. Musci, E. Stuart, A. Ruble, M. B. Beaudry, B. Schweizer, M. Owen, C. Goode, S. L. Johnson, C. Bradshaw, H. Wilcox, and K. Swartz, "The Association of School Climate, Depression Literacy, and Mental Health Stigma Among High School Students," *Journal of School Health*, vol. 87, no. 8, pp. 567–574, Aug. 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/josh.12527

[16] M. V. Birk, G. Wadley, V. V. Abeele, R. Mandryk, and J. Torous, "Video games for mental health," *Interactions*, vol. 26, no. 4, pp. 32–36, Jun. 2019. [Online]. Available: https://dl.acm.org/doi/10.1145/3328483

[17] D. Coyle, G. O' Reilly, H. van der Meulen, C. Tunney, P. Cooney, and C. Jackman, *Pesky gNATs: Using Games to Support Mental Health Interventions for Adolescents*, Oct. 2017.

[18] J. Cocks, *Synthesising Behaviour Change Theory with Game Design Practice.*, Oct. 2017.

[19] C. Bérubé, T. Schachner, R. Keller, E. Fleisch, F. V Wangenheim, F. Barata, and T. Kowatsch, "Voice-Based Conversational Agents for the Prevention and Management of Chronic and Mental Health Conditions: Systematic Literature Review," *Journal of Medical Internet Research*, vol. 23, no. 3, p. e25933, Mar. 2021.

[20] Y. He, L. Yang, C. Qian, T. Li, Z. Su, Q. Zhang, and X. Hou, "Conversational Agent Interventions for Mental Health Problems: Systematic Review and Meta-analysis of Randomized Controlled Trials," *Journal of Medical Internet Research*, vol. 25, p. e43862, Apr. 2023.

[21] A. I. Jabir, L. Martinengo, X. Lin, J. Torous, M. Subramaniam, and L. Tudor Car, "Evaluating Conversational Agents for Mental Health: Scoping Review of Outcomes and Outcome Measurement Instruments," *Journal of Medical Internet Research*, vol. 25, p. e44548, Apr. 2023. [Online]. Available: https://www.jmir.org/2023/1/e44548

[22] F. Alqahtani and R. Orji, "Insights from user reviews to improve mental health apps," *Health Informatics Journal*, vol. 26, no. 3, pp. 2042–2066, Sep. 2020. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1460458219896492

[23] N. A. J. De Witte, S. Joris, E. Van Assche, and T. Van Daele, "Technological and Digital Interventions for Mental Health and Wellbeing: An Overview of Systematic Reviews," *Frontiers in Digital Health*, vol. 3, p. 754337, Dec. 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fdgth.2021.754337/full

[24] J. Camingue, E. Carstensdottir, and E. F. Melcer, "What is a Visual Novel?" *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CHI PLAY, pp. 1–18, Oct. 2021. [Online]. Available: https://dl.acm.org/doi/10.1145/3474712

[25] M. U. Hadi, Q. Al-Tashi, R. Qureshi, A. Shah, A. Muneer, M. Irfan, A. Zafar, M. Shaikh, N. Akhtar, J. Wu, and S. Mirjalili, *Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects*, Nov. 2023.

[26] E.-y. Lee, N. G. D. Il, G.-h. An, S. Lee, and K. Lim, "ChatGPT-Based Debate Game Application Utilizing Prompt Engineering," in *Proceedings of the International Conference on Research in Adaptive and Convergent Systems*. Gdansk Poland: ACM, Aug. 2023, pp. 1–6. [Online]. Available: https://dl.acm.org/doi/10.1145/3599957.3606244

[27] V. Kumaran, J. Rowe, B. Mott, and J. Lester, "SceneCraft: Automating Interactive Narrative Scene Generation in Digital Games with Large Language Models," *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 19, no. 1, pp. 86–96, Oct. 2023. [Online]. Available: https://ojs.aaai.org/index.php/AIIDE/article/view/27504

[28] V. Nisi, S. James, P. Bala, A. Del Bue, and N. Nunes, "Inclusive Digital Storytelling: Artificial Intelligence and Augmented Reality to Re-centre Stories from the Margins," Oct. 2023, pp. 117–137.

[29] T. Fasihi Harandi, M. Mohammad Taghinasab, and T. Dehghan Nayeri, "The correlation of social support with mental health: A meta-analysis," *Electronic Physician*, vol. 9, no. 9, pp. 5212–5222, Sep. 2017. [Online]. Available: http://www.ephysician.ir/index.php/browse-issues/2017/9/802-5212

[30] J. Zimmerman, J. Forlizzi, and S. Evenson, "Research through design as a method for interaction design research in HCI," in *Proceedings of the SIGCHI Conference on Human*

Factors in Computing Systems. San Jose California USA: ACM, Apr. 2007, pp. 493–502. [Online]. Available: https://dl.acm.org/doi/10.1145/1240624.1240704

[31] J. P. Vieira Sousa, P. Campos, and P. Bala, "College Tales: Pilot Study on Large Language Models Generated Narratives for Mental Health Literacy," in *Proceedings of the 27th International Academic Mindtrek Conference*. Tampere Finland: ACM, Oct. 2024, pp. 270–275. [Online]. Available: https://dl.acm.org/doi/10.1145/3681716.3689447

[32] O. Remes, J. a. F. Mendes, and P. Templeton, "Biological, Psychological, and Social Determinants of Depression: A Review of Recent Literature," *Brain Sciences*, vol. 11, no. 12, p. 1633, Dec. 2021.

[33] A. P. Association, *Stress in America 2020: A national mental health crisis*, 2020.

[34] CDC, "Youth Risk Behavior Survey Data Summary & Trends Report: 2013-2023," Tech. Rep., Feb. 2025. [Online]. Available: https://www.cdc.gov/yrbs/dstr/index.html

[35] D. Eisenberg, S. K. Lipson, J. Heinze, and S. Zhou, "The Healthy Minds Study 2022-2023 Data Report," Tech. Rep.

[36] "WHO - Depressive disorder (depression)," publisher: World Health Organization. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[37] "American Psychological Association - Depressive Disorder," publisher: APA Dictionary of Psychology. [Online]. Available: https://dictionary.apa.org/depressive-disorder

[38] A. P. Association and A. P. Association, Eds., *Diagnostic and statistical manual of mental disorders: DSM-5*, 5th ed. Washington, D.C: American Psychiatric Association, 2013.

[39] D. Marynowski-Traczyk, M. Broadbent, S. A. Kinner, G. FitzGerald, E. Heffernan, A. Johnston, J. T. Young, G. Keijzers, P. Scuffham, E. Bosley, M. Martin-Khan, P. Zhang, and J. Crilly, "Mental health presentations to the emergency department: A perspective on the involvement of social support networks," *Australasian Emergency Care*, vol. 22, no. 3, pp. 162–167, Sep. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2588994X19300454

[40] "American Psychological Association - Support," publisher: APA Dictionary of Psychology. [Online]. Available: https://dictionary.apa.org/social-support

[41] A. F. Jorm, A. E. Korten, P. A. Jacomb, H. Christensen, B. Rodgers, and P. Pollitt, ""Mental health literacy": a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment," *Medical Journal of Australia*, vol. 166, no. 4, pp. 182–186, 1997, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.5694/j.1326-5377.1997.tb140071.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.5694/j.1326-5377.1997.tb140071.x

[42] A. F. Jorm, "Mental health literacy. Public knowledge and beliefs about mental disorders," *The British Journal of Psychiatry: The Journal of Mental Science*, vol. 177, pp. 396–401, Nov. 2000.

[43] A. Jorm, "We need to move from 'mental health literacy' to 'mental health action'," *Mental Health & Prevention*, vol. 18, p. 200179, Jun. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212657020300222

[44] P. W. Corrigan and A. C. Watson, "The paradox of self-stigma and mental illness." *Clinical Psychology: Science and Practice*, vol. 9, no. 1, pp. 35–53, 2002. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1093/clipsy.9.1.35

[45] B. G. Link and J. C. Phelan, "Conceptualizing stigma," *Annual review of Sociology*, vol. 27, no. 1, pp. 363–385, 2001, iSBN: 0360-0572 Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

[46] K. M. E. Janssens, M. C. W. Joosen, C. Henderson, J. van Weeghel, and E. P. M. Brouwers, "Improving Work Participation Outcomes Among Unemployed People with Mental Health Issues/Mental Illness: Feasibility of a Stigma Awareness Intervention," *Journal of Occupational Rehabilitation*, Oct. 2023. [Online]. Available: https://doi.org/10.1007/s10926-023-10141-3

[47] "American Psychological Association - Awareness," publisher: APA Dictionary of Psychology. [Online]. Available: https://dictionary.apa.org/awareness

[48] J. Newell, M. Gotsis, C. Lelon, and G. Davison, *Games for Psychotherapeutic Change: A Systematic Review of the Positive, Long-Term Impact of Video Game Play on Anxiety and Depression*, Oct. 2017.

[49] T. M. Li, M. Chau, P. W. Wong, E. S. Lai, and P. S. Yip, "Evaluation of a Web-Based Social Network Electronic Game in Enhancing Mental Health Literacy for Young People," *Journal of Medical Internet Research*, vol. 15, no. 5, p. e2316, May 2013, company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada. [Online]. Available: https://www.jmir.org/2013/5/e80

[50] D. Labrosse, C. Vié, H. Hajjam, C. Tisseron, D. Thellier, and I. Montagni, "An Escape Game on University Students' Mental Health During the COVID-19 Pandemic: Cocreation Study," *JMIR Serious Games*, vol. 12, no. 1, p. e48545, Mar. 2024, company: JMIR Serious Games Distributor: JMIR Serious Games Institution: JMIR Serious Games Label: JMIR Serious Games Publisher: JMIR Publications Inc., Toronto, Canada. [Online]. Available: https://games.jmir.org/2024/1/e48545

[51] A. J. Cangas, N. Navarro, J. M. Aguilar-Parra, R. Trigueros, J. Gallego, R. Zárate, and M. Gregg, "Analysis of the Usefulness of a Serious Game to Raise Awareness about Mental Health Problems in a Sample of High School and University Students: Relationship with Familiarity and Time Spent Playing Video Games," *Journal of Clinical Medicine*, vol. 8, no. 10, p. 1504, Oct. 2019, number: 10 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2077-0383/8/10/1504

[52] A. Tuijnman, M. Kleinjan, M. Olthof, E. Hoogendoorn, I. Granic, and R. C. Engels, "A Game-Based School Program for Mental Health Literacy and Stigma on Depression (Moving Stories): Cluster Randomized Controlled Trial," *JMIR Mental Health*, vol. 9, no. 8, p. e26615, Aug. 2022, company: JMIR Mental Health Distributor: JMIR Mental Health Institution: JMIR Mental Health Label: JMIR Mental Health Publisher: JMIR Publications Inc., Toronto, Canada. [Online]. Available: https://mental.jmir.org/2022/8/e26615

[53] B. Ridout and A. Campbell, "The Use of Social Networking Sites in Mental Health Interventions for Young People: Systematic Review," *Journal of Medical Internet Research*, vol. 20, no. 12, p. e12244, Dec. 2018.

[54] J. Mitchell, K. Vella, D. Johnson, N. Peever, V. Cheng, T. Davenport, J. Burns, I. Hickie, A. Kyle, and B. Hedley, "MindMax: Using videogames and sport to engage young men and

improve wellbeing," in *Proceedings of the 2nd Symposium Computing and Mental Health, 2017*. MIT Media Laboratory, 2017, pp. 1–5.

[55] N. Peever, K. Vella, D. Johnson, B. Ploderer, M. Klarkowski, and J. Mitchell, "Understanding initial experiences with Mindmax, an mHealth app that draws on shared interests in sports and video games," in *Proceedings of the 29th Australian Conference on Computer-Human Interaction*. Brisbane Queensland Australia: ACM, Nov. 2017, pp. 438–442. [Online]. Available: https://dl.acm.org/doi/10.1145/3152771.3156152

[56] D. C. Watkins, J. O. Allen, J. R. Goodwill, and B. Noel, "Strengths and weaknesses of the Young Black Men, Masculinities, and Mental Health (YBMen) Facebook project." *American Journal of Orthopsychiatry*, vol. 87, no. 4, p. 392, 2017, iSBN: 1939-0025 Publisher: Educational Publishing Foundation.

[57] W. Smith, G. Wadley, S. Webber, B. Ploderer, and R. Lederman, "Unbounding the interaction design problem: the contribution of HCI in three interventions for well-being," in *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*, 2014, pp. 392–395.

[58] S. Ito-Jaeger, E. Perez Vallejos, T. Curran, V. Spors, Y. Long, A. Liguori, M. Warwick, M. Wilson, and P. Crawford, "Digital video interventions and mental health literacy among young people: a scoping review," *Journal of Mental Health*, vol. 31, no. 6, pp. 873–883, Nov. 2022, publisher: Routledge _eprint: https://doi.org/10.1080/09638237.2021.1922642. [Online]. Available: https://doi.org/10.1080/09638237.2021.1922642

[59] S. Zhou, L. B. Hendra, Q. Zhang, J. Holopainen, and R. LC, "Eternagram: Probing Player Attitudes in Alternate Climate Scenarios Through a ChatGPT-Driven Text Adventure," 2024, version Number: 1. [Online]. Available: https://arxiv.org/abs/2403.18160

[60] B. Meskó, "Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial," *Journal of Medical Internet Research*, vol. 25, p. e50638, Oct. 2023.

[61] P. Brookins and J. M. DeBacker, "Playing Games With GPT: What Can We Learn About a Large Language Model From Canonical Strategic Games?" *SSRN Electronic Journal*, 2023. [Online]. Available: https://www.ssrn.com/abstract=4493398

[62] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. Hartshorn, S. Hosseini, R. Hou, H. Inan, M. Kardas, V. Kerkez, M. Khabsa, I. Kloumann, A. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovich, Y. Lu, Y. Mao, X. Martinet, T. Mihaylov, P. Mishra, I. Molybog, Y. Nie, A. Poulton, J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. E. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, and T. Scialom, "Llama 2: Open Foundation and Fine-Tuned Chat Models," 2023, version Number: 2. [Online]. Available: https://arxiv.org/abs/2307.09288

[63] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, Jan. 2006. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa

[64] R. C. Martin and R. C. Martin, *Clean architecture: a craftsman's guide to software structure and design*, ser. Robert C. Martin series. London, England: Prentice Hall, 2018, oCLC: on1004983973.

[65] K. Beck, *Test-driven development: by example*, ser. The Addison-Wesley signature series. Boston: Addison-Wesley, 2003.

[66] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behavior Research Methods*, vol. 39, no. 2, pp. 175–191, May 2007. [Online]. Available: http://link.springer.com/10.3758/BF03193146

[67] M. Green and T. Brock, "The Role of Transportation in the Persuasiveness of Public Narrative," *Journal of personality and social psychology*, vol. 79, pp. 701–21, Nov. 2000.

[68] K. Kulwicka and A. Gasiorowska, "Depression literacy and misconceptions scale (DepSter): a new two-factorial tool for measuring beliefs about depression," *BMC Psychiatry*, vol. 23, no. 1, p. 300, May 2023. [Online]. Available: https://bmcpsychiatry.biomedcentral.com/articles/10.1186/s12888-023-04796-8

[69] M. Marozzi, "Testing for concordance between several criteria," *Journal of Statistical Computation and Simulation*, vol. 84, no. 9, pp. 1843–1850, Sep. 2014, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00949655.2013.766189. [Online]. Available: https://doi.org/10.1080/00949655.2013.766189

[70] T. H. C. Fong and W. W. S. Mak, "The Effects of Internet-Based Storytelling Programs (Amazing Adventure Against Stigma) in Reducing Mental Illness Stigma With Mediation by Interactivity and Stigma Content: Randomized Controlled Trial," *Journal of Medical Internet Research*, vol. 24, no. 8, p. e37973, Aug. 2022, company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada. [Online]. Available: https://www.jmir.org/2022/8/e37973

[71] L. J. Hinyard and M. W. Kreuter, "Using Narrative Communication as a Tool for Health Behavior Change: A Conceptual, Theoretical, and Empirical Overview," *Health Education & Behavior*, vol. 34, no. 5, pp. 777–792, Oct. 2007. [Online]. Available: https://journals.sagepub.com/doi/10.1177/1090198106291963

A Ethics Commision Approval

UNIVERSIDADE da MADEIRA
COMISSÃO DE ÉTICA

PARECER N.º 146/CEUMA/2024, DE 17 DE OUTUBRO

Sobre o pedido de apreciação pela Comissão de Ética da UMa do projeto de investigação: *"Leveraging LLM Generated Narratives in Visual Novel Games for Mental Health Literacy."*

A – RELATÓRIO

O pedido de apreciação do processo n.º 146 de 10 de outubro de 2024 é referente ao projeto de investigação intitulado: " *Leveraging LLM Generated Narratives in Visual Novel Games for Mental Health Literacy."*, a ser desenvolvido no âmbito de um Mestrado em Engenharia Informática, tendo como Principal Investigador o mestrando **José Pedro Vieira Sousa** e os orientadores os Professores Doutores Pedro Filipe Ferreira Campos, da FCEE da Universidade da Madeira e Interactive Technologies Institute (ITI), Laboratório de Robótica e Engenharia de Sistemas (LARSyS) e Paulo Bala, do Instituto Superior Técnico, Universidade de Lisboa e ITI, LARSyS.

O pedido em análise é constituído pelos documentos dirigidos à Comissão de Ética da UMa, que integram o formulário do pedido, a saber:

a) Pedido de Parecer à CEUMa, datado de 10 de agosto de 2024;
b) Documento de informação ao participante e consentimento informado, livre e esclarecido para participação em estudos de investigação;
c) Instrumentos a ser usados para recolha de dados;
d) Parecer favorável emitido pelo Encarregado de Proteção de Dados da Universidade da Madeira;
e) A indicação do número Ciência ID ou ORCID do investigador principal e outros membros da equipe de investigação;
f) Declaração/termo de responsabilidade do orientador Prof. Doutor Paulo Alexandre Câmara Bala;

Tem data de início prevista para 18/10/2024 e términus a 30/11/2024 e respeita a um estudo empírico com os objetivos de verificar a validade de utilização de *Large Language Model*s (LLMs) para gerar narrativas para aumentar a literacia para saúde mental, com o foco em literacia para depressão, e avaliar a capacidade das narrativas geradas de criar retenção do conhecimento e aumento da literacia nos participantes.
A fundamentação teórica e justificação científica da investigação evidencia a pertinência do estudo com recurso a intervenções digitais para a promoção da literacia em saúde mental.
A população alvo será constituída por estudantes universitários e os critérios de inclusão serão:
- idade superior a 18 anos, considerando que é uma das faixas etárias em que começa a existir uma prevalência significativa de condições depressivas;
- capacidade de ler em Inglês e de interagir com um jogo móvel;
O recrutamento será realizado através de divulgação online e os participantes serão na maioria uma amostra de conveniência recrutada na Universidade da Madeira, ARDITI e ITI/LARSyS.
O tamanho da amostra necessário é de 28 participantes, calculado usando o software G*Power.
A recolha de dados integra:
Antes do Jogo:
- Questionário com questões de cariz sociodemográfico (e.g, sexo, idade, nível académico, email para envio dos questionários)

UNIVERSIDADE da MADEIRA
COMISSÃO DE ÉTICA

- Depression Literacy and Misconceptions Scale
- Os participantes irão interagir com o jogo através de um equipamento do pesquisador (telemóvel Android). Será ainda recolhida a narrativa criada no jogo, para avaliação das respostas da LLM.

Após o jogo
- Narrative Transportation Scale
- Depression Literacy and Misconceptions Scale
- Entrevista semiestruturada com gravação áudio, com distorção de voz.

Após 1 semana, os participantes irão receber um pedido para preenchimento de outro questionário (Depression Literacy and Misconceptions Scale).

É assegurada a participação voluntária, sem contrapartida de recompensação e o direito de desistir a qualquer momento.

Dependendo do participante a temática da narrativa poderá ser incomodativa; no formulário de consentimento informado, este problema é indicado e é inquirido ao participante se pretende continuar sendo ainda fornecido ao participante uma folha com contactos caso necessitem de serviços psicológicos (e.g., Serviço de Consulta Psicológica da Universidade da Madeira).

Os dados serão guardados apenas para o propósito do estudo, não sendo retido para outros fins. Os dados serão armazenados em espaço físico fechado a terceiros e/ou em formato digital em pastas de acesso restrito no Onedrive da instituição (Universidade da Madeira), apenas acessível aos membros da equipa de investigação. Os dados serão anonimizados contendo apenas um identificador para conectar os dados dos vários questionários com a narrativa. Será obtido o consentimento informado antes das sessões de recolha de dados, informando o utilizador sobre os objetivos, os procedimentos de recolha de dados, as tarefas e a forma de proteção dos dados. Nenhuma informação passível de identificação pessoal será tornada pública ou cedida a terceiros. A informação recolhida será guardada por dois anos após tornados públicos os resultados do projeto, sendo depois eliminada (físicos e/ou digitais).

Não prevê o retorno direto dos resultados ao participante. Está prevista a divulgação e disseminação de resultados na dissertação de mestrado e em outras formas de publicações científicas.

B - CONCLUSÃO

Considerando que estão contemplados os requisitos éticos fundamentais, a CEUMa deliberou, por unanimidade, emitir **parecer favorável** à realização do estudo, tendo em conta os pressupostos indicados pelo principal investigador.

Foi relatora deste parecer: Professora Mestre Clementina Morna

Aprovado, por unanimidade, em reunião do dia 17 de outubro de 2024

A Presidente da CEUMa

Professora Doutora Liliana Maria Gonçalves Rodrigues de Góis

B Pilot Study - Informed Consent

Information for Participant and Informed Consent

Study Name: Concept Validation for Large Language Models (LLM) Generated Narratives for Mental Health Literacy

Principal Investigators: José Pedro Sousa, Universidade da Madeira, 2027617@student.uma.pt
Other Investigators: Paulo Bala, ITI-LARSyS, paulo.bala@tecnico.ulisboa.pt; Pedro Campos, ITI-LARSyS;

Study Objective

The purpose of this study is to explore the interaction with a game prototype about mental health.

Procedure

- As a recruited participant, the researcher will explain the study and present you with an informed consent form.
- If presential, you will be handed a laptop computer with the game prototype. If online, through a video conferencing software, you will be given control of the researcher's screen, and will be able to interact with the game prototype.
- This prototype is a prompt-based game using a large language model; a chatbot/agent will ask you for interaction throughout the game and both input and output are shown in a text format.
- After the use of the prototype, the researcher will ask you some questions about the prototype. Audio will be recorded.

The expected duration of the study is under 20 minutes.

Inclusion Criteria

You are eligible for participation if you: are older than 18 years old; are able to read and use a computer in English/Portuguese; present no previous conditions that stop you from completing a task with a prompt-based game; are motivated to participate; and have a background/work experience in Psychology.

Risks

The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in using a computer in your daily life. The topic of the game prototype (mental health) may be sensitive, but due to your background/work experience will not likely cause discomfort.

Benefits

There may be no personal benefit from your participation in the study but the knowledge received may be of value to humanity. Particularly, it will contribute to the understanding, refinement and development of prompt-based games and experiences.

Compensation & Costs

There will be no compensation or costs to you if you participate in this study.

Confidentiality

By participating in the study, you understand and agree that we may be required to disclose your consent form, data and other personally identifiable information as required by law, regulation, subpoena or court order. Otherwise, your confidentiality will be maintained in the following manner:

- Your data and consent form will be kept separate. Your consent form will be stored in a locked location on our research institute and will not be disclosed to third parties.
- By participating, you understand and agree that the data and information gathered during this study may be used by the research team and published and/or disclosed to others outside of the

research team. However, your name, contact information and other direct personal identifiers will not be mentioned in any such publication or dissemination of the research data and/or results.

- To protect your privacy you will be assigned a code number and the collected data will be recorded by this code, not your name. This task will be done by the researcher(s) carrying out the study.
- Any audio recordings will be transcribed and identified with the participant code. The original audio recordings will be deleted. This task will be done by the researcher(s) carrying out the study.
- Only authorized researchers will have access to these data. Raw data is only accessible to researcher(s) carrying out the study. The rest of the research team only has access to data that has been anonymized as mentioned above.
- Data is kept for two years after the study is finished, being deleted after that time.

Rights

Your participation is voluntary. You are free to stop your participation at any point. Refusal to participate or withdrawal of your consent or discontinued participation in the study will not result in any penalty or loss of benefits or rights to which you might otherwise be entitled. The Principal Investigators may at his/her discretion remove you from the study for any of a number of reasons. In such an event, you will not suffer any penalty or loss of benefits or rights which you might otherwise be entitled.

Right to Ask Questions & Contact Information

If you have any questions about this study, you should feel free to ask them now. The Investigator overseeing the study will answer as long as the answer doesn't interfere with the goal of study. If you have questions later, desire additional information, or wish to withdraw your participation please contact one of the principal Investigators by e-mail in accordance with the contact information listed on the first page of this consent.

Voluntary Consent

By signing below, you agree that the above information has been explained to you and all your current questions have been answered. You are encouraged to ask questions about any aspect of this research study during the course of the study and in the future; the overseeing investigator will use his/hers discretion to answer in a way that doesn't interfere with the goals of the study. By signing this form, you agree to participate in this research study. You will receive a copy of this informed consent form signed and dated.

_____ _____

PARTICIPANT SIGNATURE DATE

Investigator obtaining Consent

As a member of the research team, I certify that I have explained the nature and purpose of this research study to the above individual and I have discussed the potential benefits and possible risks of participation in the study. Any questions the individual has about this study have been answered and any future questions will be answered as they arise.

_____ _____

SIGNATURE OF PERSON OBTAINING CONSENT DATE

C Pilot Study - Rules and Guidelines

Rules

- Rule 1 - There are two types of scenes, that are to be used:
  - Text Scenes: Present the story without choices, that continue the story until a new choice scene.
  - Choice Scenes: Offer a situation with three options for the player to choose from that can change the course of the rest of the story.
- Rule 2 - The narrative is generated based on the player's choices in the choice scenes.
- Rule 3 - Each scene is displayed individually. After each scene, the player provides the acceptable command to proceed.
- Rule 4 - The scenes feature an inner monologue by the main character, offering insight into their feelings before the player selects an option.
- Rule 5 - The player gives the main character a name and defines their gender in the first scene. Other character names are random.
- Rule 6 - Text and choices from previous scenes are saved but not displayed in subsequent scenes.
- Rule 7 - Players have four commands:
  - "Next": Progresses in text scenes.
  - "Choice $": Used in choice scenes to select an option.
  - "Show": Displays all previous scenes in order.
  - "Continue": Resumes the game after displaying previous scenes.
- Rule 8 - Scenes are sequentially numbered for reference.
- Rule 9 - The game culminates in scene 20, this is a text scene, also providing an evaluation based on the main character's story and choices: "Good Ending," "Neutral Ending," or "Bad Ending."
- Rule 10 - Scenes are formatted as follows:
  - Text Scenes:
    Scene
    [Narrative Text]
    [Inner Monologue]

  - - Choice Scenes:
    Scene
    [Narrative Text]
    [Inner Monologue]
    A - [Choice 1]
    B - [Choice 2]
    C - [Choice 3]

Guidelines

- Setting - College
- Theme – Depression
- Theme Description - The story must make use of the depression symptoms and stigmas to tell a story that can show the player that the stigmas are false and increase the player's depression literacy, without being forcefully in showing the depression theme as it also needs to create a believable story of a college student.
- Characters:
    - Main Character (MC) - will be defined the genre and name by the player with no default;
    - MC Dad
    - MC Mom
    - MC Friend 1
    - MC Friend 2
    - MC Friend 3
    - College Teacher
    - College Therapist
- Topics that can't be used:
    - Suicide

D Pilot Study - Semi-Structured Interview

Questions

- About the prototype:

    1. Was the prototype effective in presenting depression and the challenges associated with it, including symptoms identified in DSM-5?

        - (followup) How? Or why was is not effective?

    2. Was the prototype effective in addressing the stigma surrounding depression?

        - (followup) How? Or why was is not effective?

    3. Was the story believable regarding the topic of depression?

        - (followup) How? Or why was is not effective?

- About their practice:

    4. Based on your clinical experience/training, what strategies or techniques would you apply to increase awareness and literacy about depression among patients?

    5. Based on your clinical experience/training, what strategies or techniques would you apply to build/encourage support networks for patients?

- About the rules/guidelines:

    (Give them the rules and guidelines, ask them to read them.)

    6. Given your play experience and after reading them, how do you perceive the guidelines and rules given for the narrative?

        - (followup) Were there rules/guidelines that you feel were ignored? Is there rules/guidelines that should be present?

    7. What recommendations would you make for refining or expanding the guidelines/rules?

Table 16: Code Groups Pilot Study

Code Group	Initial Codes	Rationale
CG1 - Missing Information	C3 - Missing Symptoms	
	C14 - Missing Information	
	C15 - Need for Context	There is a need for information about context, and background information, thus missing.
	C32 - Ineffective Prototype	Ineffective because doesn't identify the criteria, thus missing that info.
	C33 - Missing Criteria	
CG2 - Simplification of representation of Depression	C1 - Oversimplification of Depression	Simplification of depression (in general).
	C9 - Misinformation	Lack of information, about going to psychologists (it borders stigma).
	C11 - Daily life Representation	Shows the daily life of a person with depression (opposed view of the predominant view, thus included in this theme).
	C13 - Narrative Depth for small slices of life	More depth on the narrative, because it focus on small slices of life (can be considered as simplifying because doesn't show that much information on the reality of depression).
	C22 - Students Reality	Presents the student's reality, as it's the intended target group of the study to be developed considering the results of this pilot, thus it was included in this grouped code as it is a view, that implies a representation of part of the life of the target group.
	C28 - Good Representation of Depression	Identified that the narrative gave a good representation of depression even if there is a simplification of the reality so it can be put in a narrative.
	C29 - No Representation of Stigma	No representation of the stigma, or that the prototype wasn't "evaluating" it.
CG3 - Representation of Recovery	C2 - Misrepresentation of Recovery	Doesn't showed the recovery as in the real world
	C4 - Recovery Reality	presents the reality of recovery (opposed view to the general trend).

Continued on next page

Table 16 – Code Groups Pilot Study

Code Group	Initial Codes	Rationale
	C5 - Role of Medication	Sometimes the need for medication and the need for a double approach, through psychology and psychiatry.
	C17 - Highs and Lows of Depression	The wrong view that depression and recovery are linear, where there are setbacks and comebacks.
	C20 - Need to give Solutions	The need to give tips and solutions, about depression and also explaining what could be done in certain situations or for certain choices.
	C25 - Journaling as Coping	Was given as way to cope, and its correct but there is also a need to explain that it doesn't work alone and on the first time and its not a cure.
	C30 - Good Representation of Internal Dialogue	The inner dialogue was right, it looked like what someone with depression can think (opposed view).
CG4 - Support Network & Narrative Problems	C10 - Support Network	Information related to the importance of support networks and their impact, e.g. "even the person who, that sometimes triggers these kind of things, can help".
	C18 - Strategies for Support	Some strategies that the therapists use.
	C19 - Narrative Looping	Narrative situations where. there is aloop in the story and it doesn't change or advance from a certain place or situation.
	C23 - Personal Connection	Connection with the narrative themes.
	C24 - Seeking Help	Narrative gives the choice to seek for help.
	C26 - Need more Choices	Could have more choices to increase the interactivity, and need more in the coping, for example in the journaling (instead of one time thing and its good).
	C31 - Importance of Physiological Perspective	Strategy of explaining the physiological perspective to understand.
CG5 - Literacy and Therapists	C6 - Need for PsychoEducation	The need to give education about psychology and the surrounding themes, like the work of a psychologist.

Continued on next page

Table 16 – Code Groups Pilot Study

Code Group	Initial Codes	Rationale
	C7 - Stigma Surrounding Mental Health	The representation of the stigma in the narrative and the choices to avoid/combat it, also the reality that there still exists stigma associated to psychologists and mental health.
	C8 - Need for Literacy	
	C12 - Importance of Structure for Professional and Patient	
	C16 - Stigma and Help Seeking	Information regarding stigma and that its okay to look for help.
	C21 - Useful Therapists	These types of prototypes and narratives can be useful for therapists.
	C27 - Role and Future of AI in Therapy	
	C34 - Importance of Clinical Understanding	Some narratives induced a doubt if it was really depression.

Table 17: Themes and Code Groups

Theme	Code Group
Missing Information	CG1 - Missing Information
Oversimplifying depression	CG2 - Simplification of representation of Depression
Misrepresenting Recovery	CG3 - Representation of Recovery
Impact of Support Networks in the Narrative	CG4 - Support Network & Narrative Problems
Potential for Mental Health Literacy	CG5 - Literacy and Therapists

E Prototype Study - Informed Consent

Consentimento Informado, Esclarecido e Livre para Participação em estudos de Investigação

Identificação do Investigador: José Pedro Sousa, Universidade da Madeira, 2027617@student.uma.pt; Paulo Bala, ITI-LARSyS, IST, paulo.bala@tecnico.ulisboa.pt; Pedro Campos, Universidade da Madeira, ITI-LARSyS;

Título do estudo: Leveraging LLM Generated Narratives in Visual Novel Games for Mental Health Literacy.

Enquadramento: Foi convidado a participar num projeto de mestrado. Com o objectivo de verificar/avaliar a validade de utilização de LLMs (Large Language Models - Modelos de linguagem de grande escala) para gerar narrativas para aumentar a literacia e verificar/avaliar a validade de utilização de LLMs para gerar narrativas sobre temas em áreas da saúde.

Explicação do estudo: No presente estudo, irá interagir com um protótipo que usa L.u;,LLMs para a criação de narrativas interactivas sobre tópicos na área da saúde. O objectivo do estudo é a avaliação do protótipo, e não dos participantes.

Antes de interagir com o protótipo, como participante terá de preencher um questionário sociodemográfico e um questionário de literacia e falsos conceitos sobre depressão (Depression Literacy and Misconceptions Scale).

Após interagir com o protótipo, como participante terá de preencher um questionário de transporte da narrativa (Narrative Transportation Scale), um questionário de literacia e falsos conceitos sobre depressão (Depression Literacy and Misconceptions Scale) e será feita uma entrevista semi-estruturada (com gravação de audio com uso de aplicação de distorção de voz).

Posteriormente a esta sessão (uma semana depois), como participante irá ser contactado para preencher um questionário de literacia e falsos conceitos sobre depressão (Depression Literacy and Misconceptions Scale).

A informação recolhida será os dados quantitativos dos questionários, os dados qualitativos da entrevista semi-estruturada e a narrativa obtida para avaliação das respostas da LLM.

Os dados serão anonimizados com um identificador, que será apenas usado para conectar os dados dos vários questionários com a narrativa, sendo destruídos dois anos após tornados públicos os resultados do estudo.

Condições e financiamento: A sua participação é voluntária, e sem contrapartida ou recompensas, pode terminar a sua participação em qualquer altura. Recusa em participar, retirada do seu consentimento ou interrupção da participação no estudo não resultará em qualquer penalidade ou perda de benefícios ou direitos aos quais você poderia ter direito. Os investigadores principais poderão, a seu critério, removê-lo do estudo por uma série de razões. Nesse caso, você não sofre nenhuma penalidade ou perda de benefícios ou direitos aos quais você poderia ter direito. Se você tiver alguma dúvida sobre este estudo, desejar informações adicionais ou desejar retirar sua participação, entre em contato com o investigador principal por e-mail, de acordo com as informações de contato listadas na segunda página deste consentimento. Este estudo foi financiado pela bolsa de Investigação do investigador principal. Este estudo foi aprovado pela comissão de ética da Universidade da Madeira.

O risco e descomforto associados à participação neste estudo serão os mesmos associados aos encontrados na vida diária. O tópico da narrativa poderá ser um tópico sensível para determinados participantes, sendo fornecido ao

participante uma folha com contactos caso necessitem de serviços psicológicos, contendo os contactos do Serviço de Psicologia da Universidade da Madeira (email: servico.psicologia@mail.uma.pt e contactos: 291 705 084 | +351 91 81 59 467).

Pode não haver nenhum benefício pessoal da sua participação no estudo além do possível aumento da literacia em saúde mental, mas o conhecimento recebido pode ser de valor para a humanidade. Em particular, contribuirá para a compreensão, aperfeiçoamento e desenvolvimento de jogos e experiências com narrativas geradas por LLMs para tópicos de saúde.

Anonimato e confidencialidade: Ao participar do estudo, entende e concorda que o projeto pode ser obrigado a divulgar o seu formulário de consentimento, dados e outras informações de identificação pessoal, conforme exigido por lei, regulamento, intimação ou ordem judicial. Caso contrário, a sua confidencialidade será mantida da seguinte maneira. Os dados e informações recolhidas durante este estudo serão anonimizados e podem ser usados e publicados e / ou divulgados pelo projeto para fins de pesquisa e/ou publicações académicas. No entanto, as suas informações pessoais associadas ao código identificador nunca serão reveladas em nenhuma publicação ou divulgação dos dados e/ou resultados da investigação.

Por favor, leia com atenção esta informação. Se achar que algo está incorreto ou que não está claro, não hesite em solicitar mais informações.

José Pedro Sousa, Universidade da Madeira, 2027617@student.uma.pt, josepedrovsousa@gmail.com;

Paulo Bala, ITI-LARSyS, paulo.bala@tecnico.ulisboa.pt;

Pedro Campos, Universidade da Madeira;

Se concorda com a proposta que lhe foi feita, queira assinar este documento.

Assinatura de quem pede consentimento: _____

UNIVERSIDADE da MADEIRA

Declaração de Consentimento do Participante

Eu, _____

declaro ter lido e compreendido este documento, bem como as informações verbais que me foram fornecidas pela/s pessoa/s que acima assina/m. Foi-me garantida a possibilidade de, em qualquer altura, recusar participar neste estudo sem qualquer tipo de consequências. Desta forma, aceito participar neste estudo e permito a utilização dos dados, que de forma voluntária forneço, confiando em que apenas serão utilizados para fins científicos e publicações que delas decorram e com as garantias de confidencialidade e anonimato que me são dadas pelo/a investigador/a.

Email (opcional): _____

Assinatura legível e manuscrita: _____

Data: ___ / ___ / _____

F Prototype Study - Semi-Structured Interview Questions

Semi-Structured Interview Guide

1. Have you ever used digital tools or applications for mental health education or support? If yes, can you tell which ones and what your experience was?

2. Did the narrative seem relatable or authentic to you?

3. Did you learn anything new about depression from the narrative? If so, what did you learn?

4. Was the information presented in the narrative clear and easy to understand?

5. Do you think the narrative could help reduce the stigma associated with depression?

6. How did the narrative make you feel? Did it evoke any particular emotions or reflections?

7. Did the narrative provide any sense of empathy or understanding towards individuals with depression?

8. Do you think the information from the narrative could be useful in real-life situations? For example, in supporting someone with depression or seeking help for yourself?

9. Would you recommend this type of narrative to others who want to learn about depression?

10. Is there anything else you would like to share about your experience with the narrative or any additional thoughts on using LLMs for mental health education?

G Prototype - Narrative Transportation Scale

Green & Brock Narrative Transportation Scale (2000)*

Narrative Questionnaire

Circle the number under each question that best represents your opinion about the narrative you just read.

1. While I was reading the narrative, I could easily picture the events in it taking place.

1	2	3	4	5	6	7
not at all						very much

2. While I was reading the narrative, activity going on in the room around me was on my mind.

1	2	3	4	5	6	7
not at all						very much

3. I could picture myself in the scene of the events described in the narrative.

1	2	3	4	5	6	7
not at all						very much

4. I was mentally involved in the narrative while reading it.

1	2	3	4	5	6	7
not at all						very much

5. After the narrative ended, I found it easy to put it out of my mind.

1	2	3	4	5	6	7
not at all						very much

6. I wanted to learn how the narrative ended.

1	2	3	4	5	6	7
not at all						very much

7. The narrative affected me emotionally.

1	2	3	4	5	6	7
not at all						very much

8. I found myself thinking of ways the narrative could have turned out differently.

1	2	3	4	5	6	7
not at all						very much

9. I found my mind wandering while reading the narrative.

1	2	3	4	5	6	7
not at all						very much

10. The events in the narrative are relevant to my everyday life.

1	2	3	4	5	6	7
not at all						very much

11. The events in the narrative have changed my life.

1	2	3	4	5	6	7
not at all						very much

12. I had a vivid mental image of [character name].

1	2	3	4	5	6	7
not at all						very much

Notes: Items 2, 5, and 9 are reverse-scored.
Item 12 can be repeated for the number of main characters in the story, substituting a different character name for each item.

*-Citation: Green, M.C., & Brock, T.C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology, 79(5),* 701-721.

H Prototype - Depression Literacy and Misconceptions Scale

Appendix

Depression Literacy and Misconceptions Scale (DepSter)

Please Indicate Your Agreement with The Following Statements Using the Scale Below

1	2	3	4	5
Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

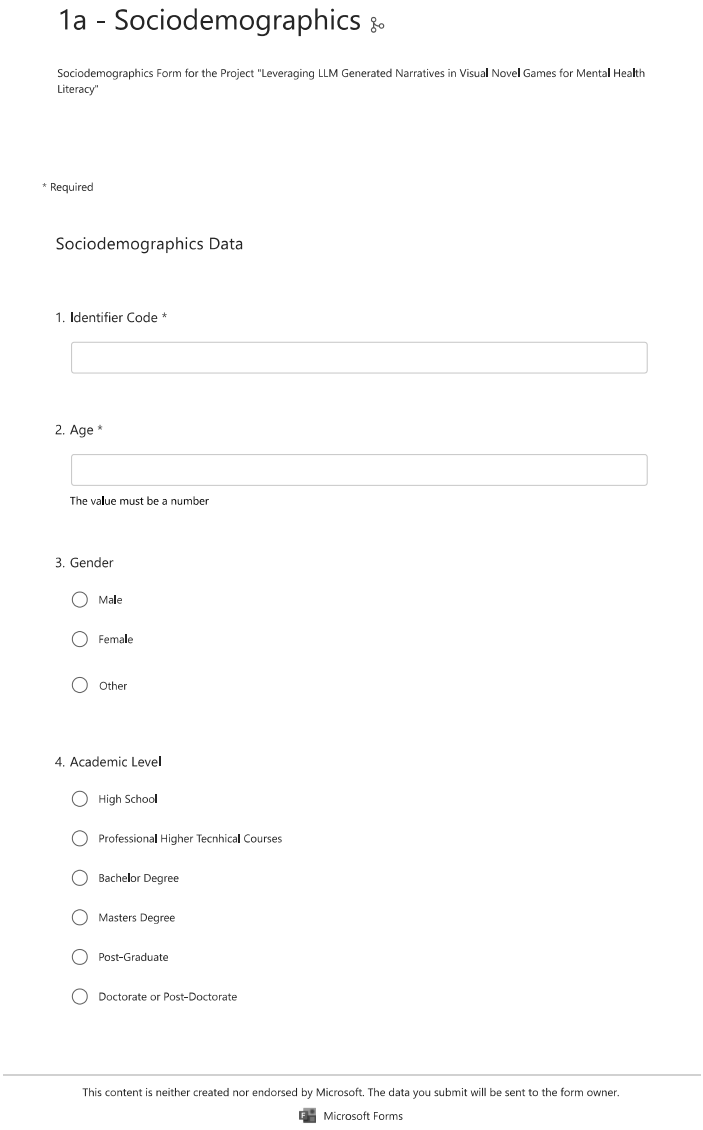
1) Depression is an illness
2) Depression is just a fad
3) Depression can affect anyone
4) Depression is just a temporary mood deterioration
5) Depression makes people lose interest even in the things they used to enjoy doing
6) Depression affects only the people who are weak and cannot cope with their life
7) Depression makes people lack the strength to do anything
8) To overcome depression, all you need is willpower
9) People with depression often think about suicide
10) To overcome depression, all you need is to get yourself together
11) Depression is associated with great suffering
12) Antidepressant medication start to work right after the intake
13) Depression is just a self-pitty
14) People with depression are mentally weak

Scoring:

Depression Literacy (DL): Average of 1, 3, 5, 7, 9, 11

Misconceptions About Depression (MiscD): Average 2, 4, 6, 8, 10, 12, 13, 14

I Prototype - Pre-test Sociodemographics Questionnaire



1a - Sociodemographics

Sociodemographics Form for the Project "Leveraging LLM Generated Narratives in Visual Novel Games for Mental Health Literacy"

* Required

Sociodemographics Data

1. Identifier Code *

2. Age *

The value must be a number

3. Gender

- Male
- Female
- Other

4. Academic Level

- High School
- Professional Higher Tecnhical Courses
- Bachelor Degree
- Masters Degree
- Post-Graduate
- Doctorate or Post-Doctorate

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

Microsoft Forms

J Prototype - Pre-Test Scale Questionnaire

1b - Depression Literacy and Misconceptions Scale

Depression Literacy and Misconceptions Scale for the Project "Leveraging LLM Generated Narratives in Visual Novel Games for Mental Health Literacy"

* Required

Depression Literacy and Misconceptions Scale (Kulwicka et al., 2023)

1. Identifier Code *

2. Depression Literacy and Misconceptions Scale (Kulwicka et al., 2023)
Please Indicate Your Agreement with The Following Statements Using the Scale Below:

	1 - Strongly disagree	2 - Disagree	3 - Neutral	4 - Agree	5 - Strongly agree
Depression is an illness	◯	◯	◯	◯	◯
Depression is just a fad	◯	◯	◯	◯	◯
Depression can affect anyone	◯	◯	◯	◯	◯
Depression is just a temporary mood deterioration	◯	◯	◯	◯	◯
Depression makes people lose interest even in the things they used to enjoy doing	◯	◯	◯	◯	◯
Depression affects only the people who are weak and cannot cope with their life	◯	◯	◯	◯	◯
Depression makes people lack the strength to do anything	◯	◯	◯	◯	◯
To overcome depression, all you need is willpower	◯	◯	◯	◯	◯
People with depression often think about suicide	◯	◯	◯	◯	◯
To overcome depression, all you need is to get yourself together	◯	◯	◯	◯	◯
Depression is associated with great suffering	◯	◯	◯	◯	◯
Antidepressant medication start to work right after the intake	◯	◯	◯	◯	◯
Depression is just a self-pitty	◯	◯	◯	◯	◯
People with depression are mentally weak	◯	◯	◯	◯	◯

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

K Prototype - Post-test Questionnaire

2 - Scales

Narrative Transportation Scale and Depression Literacy and Misconceptions Scale for the Project "Leveraging LLM Generated Narratives in Visual Novel Games for Mental Health Literacy"

* Required

Scales

1. Identifier Code *

2. Narrative Transportation Scale (Green, M.C., & Brock, T.C., 2000) *

Please indicate your agreement with the following statements, that best represents your opinion about the narrative you just read, using the scale below:

	1 – Not at all	2	3	4	5	6	7 – Very much
While I was reading the narrative, I could easily picture the events in it taking place.	◯	◯	◯	◯	◯	◯	◯
While I was reading the narrative, activity going on in the room around me was on my mind.	◯	◯	◯	◯	◯	◯	◯
I could picture myself in the scene of the events described in the narrative.	◯	◯	◯	◯	◯	◯	◯
I was mentally involved in the narrative while reading it.	◯	◯	◯	◯	◯	◯	◯
After the narrative ended, I found it easy to put it out of my mind.	◯	◯	◯	◯	◯	◯	◯
I wanted to learn how the narrative ended.	◯	◯	◯	◯	◯	◯	◯
The narrative affected me emotionally.	◯	◯	◯	◯	◯	◯	◯
I found myself thinking of ways the narrative could have turned out differently.	◯	◯	◯	◯	◯	◯	◯
I found my mind wandering while reading the narrative.	◯	◯	◯	◯	◯	◯	◯
The events in the narrative are relevant to my everyday life.	◯	◯	◯	◯	◯	◯	◯
The events in the narrative have changed my life.	◯	◯	◯	◯	◯	◯	◯
I had a vivid mental image of the Main Character [character name].	◯	◯	◯	◯	◯	◯	◯

3. Depression Literacy and Misconceptions Scale (Kulwicka et al., 2023) *

Please Indicate Your Agreement with The Following Statements Using the Scale Below:

	1 - Strongly disagree	2 - Disagree	3 - Neutral	4 - Agree	5 - Strongly agree
Depression is an illness	◯	◯	◯	◯	◯
Depression is just a fad	◯	◯	◯	◯	◯
Depression can affect anyone	◯	◯	◯	◯	◯
Depression is just a temporary mood deterioration	◯	◯	◯	◯	◯
Depression makes people lose interest even in the things they used to enjoy doing	◯	◯	◯	◯	◯
Depression affects only the people who are weak and cannot cope with their life	◯	◯	◯	◯	◯
Depression makes people lack the strength to do anything	◯	◯	◯	◯	◯
To overcome depression, all you need is willpower	◯	◯	◯	◯	◯
People with depression often think about suicide	◯	◯	◯	◯	◯
To overcome depression, all you need is to get yourself together	◯	◯	◯	◯	◯
Depression is associated with great suffering	◯	◯	◯	◯	◯
Antidepressant medication start to work right after the intake	◯	◯	◯	◯	◯
Depression is just a self-pitty	◯	◯	◯	◯	◯
People with depression are mentally weak	◯	◯	◯	◯	◯

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

Microsoft Forms

L Prototype - One Week Follow-up Questionnaire

3 - Depression Literacy and Misconceptions Scale

Depression Literacy and Misconceptions Scale for the Project "Leveraging LLM Generated Narratives in Visual Novel Games for Mental Health Literacy"

* Required

Depression Literacy and Misconceptions Scale (Kulwicka et al., 2023)

1. Identifier Code *

2. Depression Literacy and Misconceptions Scale (Kulwicka et al., 2023)
Please Indicate Your Agreement with The Following Statements Using the Scale Below:

	1 - Strongly disagree	2 - Disagree	3 - Neutral	4 - Agree	5 - Strongly agree
Depression is an illness	◯	◯	◯	◯	◯
Depression is just a fad	◯	◯	◯	◯	◯
Depression can affect anyone	◯	◯	◯	◯	◯
Depression is just a temporary mood deterioration	◯	◯	◯	◯	◯
Depression makes people lose interest even in the things they used to enjoy doing	◯	◯	◯	◯	◯
Depression affects only the people who are weak and cannot cope with their life	◯	◯	◯	◯	◯
Depression makes people lack the strength to do anything	◯	◯	◯	◯	◯
To overcome depression, all you need is willpower	◯	◯	◯	◯	◯
People with depression often think about suicide	◯	◯	◯	◯	◯
To overcome depression, all you need is to get yourself together	◯	◯	◯	◯	◯
Depression is associated with great suffering	◯	◯	◯	◯	◯
Antidepressant medication start to work right after the intake	◯	◯	◯	◯	◯
Depression is just a self-pitty	◯	◯	◯	◯	◯
People with depression are mentally weak	◯	◯	◯	◯	◯

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

M Prototype Study - Prompts

```
{
    "type": "json_schema",
    "json_schema": {
        "name": "scene",
        "strict": true,
        "schema": {
            "type": "object",
            "properties": {
                "scene": {
                    "type": "object",
                    "properties": {
                        "scene_num": {
                            "type": "number",
                            "description": "The number of the scene between 1 and
                            ↪  20."
                        },
                        "scene_type": {
                            "type": "string",
                            "enum": [
                                "Text",
                                "Choice",
                                "End"
                            ],
                            "description": "The type of the scene."
                        },
                        "scene_text": {
                            "type": "string",
                            "description": "The narrative text of the scene."
                        },
                        "scene_inner_monologue": {
                            "type": "string",
                            "description": "The inner monologue in the scene."
                        },
                        "scene_choices": {
                            "type": "object",
                            "properties": {
                                "Choice A": {
                                    "type": "string",
                                    "description": "Text for choice A if the
                                    ↪  scene_type is Choice otherwise empty
                                    ↪  string."
                                },
                                "Choice B": {
```

```
                    "type": "string",
                    "description": "Text for choice B if the
                    ↪  scene_type is Choice otherwise empty
                    ↪  string."
                },
                "Choice C": {
                    "type": "string",
                    "description": "Text for choice C if the
                    ↪  scene_type is Choice otherwise empty
                    ↪  string."
                }
            },
            "additionalProperties": false,
            "required": [
                "Choice A",
                "Choice B",
                "Choice C"
            ],
            "description": "Choices available in the scene."
        },
        "evaluation_text": {
            "type": "string",
            "description": "The evaluation text of the narrative,
            ↪  considering the choices and story of the main
            ↪  character, if the scene_type is End otherwise
            ↪  empty string."
        },
        "evaluation_value": {
            "type": "string",
            "enum": [
                "Good Ending",
                "Neutral Ending",
                "Bad Ending"
            ],
            "description": "The evaluation of the narrative, if
            ↪  the scene_type is End otherwise empty string."
        },
        "image": {
            "type": "string",
            "enum": [
                "dormSingleRoom",
                "campusEntrance",
                "collegeCorridors",
                "classroom",
                "canteen",
                "counsellorRoom",
```

```
                                "library",
                                "sportsField",
                                "park",
                                "gym",
                                "party"
                            ],
                            "description": "The name of an image to be used in
                            ↪  this scene, the image should be related to the
                            ↪  scene_text."
                        }
                    },
                    "required": [
                        "scene_num",
                        "scene_type",
                        "scene_text",
                        "scene_inner_monologue",
                        "scene_choices",
                        "evaluation_text",
                        "evaluation_value",
                        "image"
                    ],
                    "additionalProperties": false
                }
            },
            "required": [
                "scene"
            ],
            "additionalProperties": false
        }
    }
}
```

Prompt 9: Scenes and End Scene Schema

```
{
    "type": "json_schema",
    "json_schema": {
        "name": "scene",
        "strict": true,
        "schema": {
            "type": "object",
            "properties": {
                "scene": {
                    "type": "object",
                    "properties": {
                        "Title": {
                            "type": "string",
                            "description": "Title of the narrative."
                        },
                        "Setting": {
                            "type": "string",
                            "description": "The setting of the narrative."
                        },
                        "Characters": {
                            "type": "object",
                            "properties": {
                                "MC Dad": {
                                    "type": "object",
                                    "properties": {
                                        "Name": {
                                            "type": "string",
                                            "description": "Name of MC Dad."
                                        },
                                        "Personality": {
                                            "type": "string",
                                            "description": "Description of MC
                                            ↪  Dad's Personality."
                                        }
                                    },
                                    "required": [
                                        "Name",
                                        "Personality"
                                    ],
                                    "additionalProperties": false
                                },
                                "MC Mom": {
                                    "type": "object",
                                    "properties": {
                                        "Name": {
                                            "type": "string",
```

```
                "description": "Name of MC Mom."
            },
            "Personality": {
                "type": "string",
                "description": "Description of MC
                ↪  Mom's Personality."
            }
        },
        "required": [
            "Name",
            "Personality"
        ],
        "additionalProperties": false
    },
    "MC Friend 1": {
        "type": "object",
        "properties": {
            "Name": {
                "type": "string",
                "description": "Name of MC Friend 1."
            },
            "Personality": {
                "type": "string",
                "description": "Description of MC
                ↪  Friend 1's Personality."
            }
        },
        "required": [
            "Name",
            "Personality"
        ],
        "additionalProperties": false
    },
    "MC Friend 2": {
        "type": "object",
        "properties": {
            "Name": {
                "type": "string",
                "description": "Name of MC Friend 2."
            },
            "Personality": {
                "type": "string",
                "description": "Description of MC
                ↪  Friend 2's Personality."
            }
        },
```

```
            "required": [
                "Name",
                "Personality"
            ],
            "additionalProperties": false
        },
        "MC Friend 3": {
            "type": "object",
            "properties": {
                "Name": {
                    "type": "string",
                    "description": "Name of MC Friend 3."
                },
                "Personality": {
                    "type": "string",
                    "description": "Description of MC
                    ↪  Friend 3's Personality."
                }
            },
            "required": [
                "Name",
                "Personality"
            ],
            "additionalProperties": false
        },
        "College Teacher": {
            "type": "object",
            "properties": {
                "Name": {
                    "type": "string",
                    "description": "Name of College
                    ↪  Teacher."
                },
                "Personality": {
                    "type": "string",
                    "description": "Description of
                    ↪  College Teacher's Personality."
                }
            },
            "required": [
                "Name",
                "Personality"
            ],
            "additionalProperties": false
        },
        "College Therapist": {
```

```
                                "type": "object",
                                "properties": {
                                    "Name": {
                                        "type": "string",
                                        "description": "Name of College
                                        ↪  Therapist."
                                    },
                                    "Personality": {
                                        "type": "string",
                                        "description": "Description of
                                        ↪  College Therapist's Personality."
                                    }
                                },
                                "required": [
                                    "Name",
                                    "Personality"
                                ],
                                "additionalProperties": false
                            }
                        },
                        "required": [
                            "MC Dad",
                            "MC Mom",
                            "MC Friend 1",
                            "MC Friend 2",
                            "MC Friend 3",
                            "College Teacher",
                            "College Therapist"
                        ],
                        "description": "Characters in the narrative.",
                        "additionalProperties": false
                    }
                },
                "required": [
                    "Title",
                    "Setting",
                    "Characters"
                ],
                "additionalProperties": false
            }
        },
        "required": [
            "scene"
        ],
        "additionalProperties": false
    }
```

```
        }
    }
```

Prompt 10: Synopse Schema

```
{
    "role": "system",
    "content":  "Guidelines for the narrative:
                1. Types of Scenes:
                    - Synopse Scene: Contains the synopsis and information about some of the characters.
                    - Text Scenes: Continue the story without choices.
                    - Choice Scenes: Offer a situation with three options for the player to choose from, altering the course of the story. The choices need to lead to different outcomes, for example, a neutral, a good, and a bad one.
                    - End Scene: The final scene, providing an evaluation of the narrative, considering the player choices (Good, Neutral, or Bad Ending), while also giving options on how to improve mental health outcomes.
                2. Craft branching narratives based on the choices in choice scenes.
                3. Don't force the narrative in a specific direction - provide various options for the player to steer the story.
                4. Avoid a cyclical narrative where the character remains stuck in one location unless the player makes a specific choice.
                5. Each scene should include an inner monologue offering insight into the main character's emotions before making a choice or transitioning to the next scene.
                6. Include at least 2 natural interactions between characters in the narrative, beyond the player's direct choices. For example, even if the main character doesn't initiate a conversation, other characters may reach out through SMS, calls, or in-person dialogue. These interactions, whether welcomed or not, can have both positive and negative impacts on the main character's mental health, emphasizing the unpredictability of social relationships.
                7. Depict both positive and negative experiences for the main character, reflecting the relapsing nature of mental health recovery.
                8. Images for the scenes should depict the place where the main character is.
                9. Number all scenes sequentially for easy reference.
                10. The narrative ends at Scene 20, which is the only End Scene.
                11. Ensure that there are exactly 7 Choice Scenes, 12 Text Scenes, and 1 End Scene.
                12. Narrative Settings:
                    - Setting: College
                        - Theme: Depression
```

```
22                      - Theme Description: The story should depict symptoms of
                        ↪ depression and dispel stigmas surrounding mental
                        ↪ health, with the aim of improving readers'
                        ↪ understanding of depression. This theme should be
                        ↪ integrated seamlessly into a realistic portrayal of a
                        ↪ college student's life.
23                  - Characters:
24                      - Main Character (MC)
25                      - MC's Dad
26                      - MC's Mom
27                      - MC's Friends (1, 2, and 3)
28                      - College Teacher
29                      - College Therapist
30                  - Topics to Avoid:
31                      - Suicide
32              13. Commands:
33                  - When you receive a prompt with \"Start\" after giving the
                    ↪ settings and characters, begin the narrative. \"Start\"
                    ↪ can only be used when the previous response is the
                    ↪ Synopsis.
34                  - When you receive a prompt with \"Next\", only applicable
                    ↪ when the previous response is a Text Scene, advance to
                    ↪ the next Scene.
35                  - When you receive a prompt with \"Choice Y\" (with Y being
                    ↪ the choice), only applicable when the previous response
                    ↪ is a Choice Scene, give the next Scene following that
                    ↪ choice."
36  }
```

Prompt 11: System Guidelines Prompt

```
{
    "role": "user",
    "content": "Below are three examples, one for a text scene, one for a choice scene and one for the end scene.
        - Text Scene Example:
            ```json
                {
                    \"scene\": [
                        {
                            \"scene_num\": \"Scene 6\",
                            \"scene_type\": \"Text\",
                            \"scene_text\": \"You're sitting in your dorm room, textbooks scattered across your desk. Despite the stillness of the evening, your mind feels anything but calm. The same heavy thoughts keep looping, like a weight that refuses to lift. You try to focus, but the fatigue is more than physical-it's emotional.\",
                            \"scene_inner_monologue\": \"I've been feeling drained for weeks. No matter how much I sleep, I can't shake this exhaustion. It's like my brain is in a fog. They say depression can make everything feel harder, and maybe that's what's happening. It's not just sadness. It's this... emptiness, and it's making it impossible to care about anything.\",
                            \"images\" : [
                                \"dormRoom\",
                            ],
                        }
                    ]
                }
            ```
        - Choice Scene Example:
            ```json
                {
                    \"scene\": [
                        {
                            \"scene_num\": \"Scene 10\",
                            \"scene_type\": \"Choice\",
```
```

```
                        \"scene_text\": \"You receive a text from your friend, Sarah, asking if you want to grab dinner together. The thought of socializing feels overwhelming, but at the same time, you wonder if being around others might help. Sarah's message reads, 'Hey! Haven't seen you in a while. Want to grab dinner tonight? I'm thinking of hitting that new spot on campus.'\",
                        \"scene_inner_monologue\": \"I haven't seen anyone in days. The idea of leaving my room makes me anxious, but they say isolation can make depression worse. Maybe I should push myself to go. But what if it's just too much? It's so hard to tell if I'll feel better or worse afterward.\",
                        \"scene_choices\": {
                            \"Choice A\": \"Agree to meet Sarah for dinner.\",
                            \"Choice B\": \"Politely decline and stay in your room.\",
                            \"Choice C\": \"Suggest grabbing dinner another day when you feel more up to it.\"
                        },
                        \"images\": [
                            \"dormRoom\",
                        ],
                    }
                ]
            }
        ```
    - End Scene Examples:
        - Good Ending:
            ```json
                {
                    \"scene\": [
                        {
                            \"scene_num\": \"Scene 20\",
                            \"scene_type\": \"End\",
```

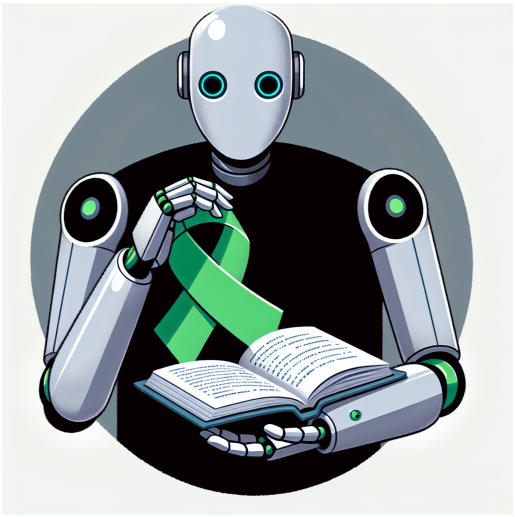
```
                            \"scene_text\": \"As you approach the end of your college journey, you take a moment to reflect on the path you've taken. Despite the hurdles along the way, you’ve built a solid foundation of resilience, self-awareness, and meaningful connections. You feel ready to embrace what comes next, knowing that you’ve learned how to care for both your mental and emotional well-being.\",
                            \"scene_inner_monologue\": \"Looking back, I realize how far I've come. There were moments when it felt like the weight of everything would never lift, but step by step, I’ve learned to reach out and let others in. I’m not 'fixed'-but I’m stronger and better equipped to handle the hard days.\",
                            \"evaluation_text\": \"Your willingness to seek help and maintain a strong support system has been instrumental in your growth. You’ve faced your challenges head-on, showing resilience and the courage to prioritize self-care. This ending reflects a journey that wasn’t without struggles, but it shows that through perseverance, meaningful progress is possible. Your experience highlights the importance of both personal effort and the support of others.\",
                            \"evaluation_value\": \"Good Ending\",
                            \"images\": [
                                \"campusEntrance\",
                            ],
                        }
                    ]
                }
            ```
        - Neutral Ending:
            ```json
                {
                    \"scene\": [
                        {
                            \"scene_num\": \"Scene 20\",
                            \"scene_type\": \"End\",
```

```
                        \"scene_text\": \"John's decision to take time
                        ↪  for himself and reflect on his feelings has
                        ↪  led to a moment of clarity and empowerment.
                        ↪  While he still faces challenges, he is
                        ↪  learning to navigate his mental health with
                        ↪  more awareness and self-compassion. This
                        ↪  journey is ongoing, but he is beginning to
                        ↪  understand the importance of self-care and
                        ↪  reaching out for support when needed.\",
                        \"scene_inner_monologue\": \"It's been a
                        ↪  challenging but enlightening journey. I've
                        ↪  learned that progress isn't always linear,
                        ↪  and that's okay. I'm not where I thought I'd
                        ↪  be, but I'm in a better place to understand
                        ↪  what I need to keep moving forward.\",
                        \"evaluation_text\": \"You've managed to balance
                        ↪  your mental health journey with personal
                        ↪  commitments, making slow but steady progress.
                        ↪  While things aren't perfect, you've shown
                        ↪  resilience by taking steps to understand and
                        ↪  care for yourself. This balanced approach
                        ↪  reflects a realistic view of mental health
                        ↪  recovery-ongoing, with both ups and downs,
                        ↪  but with moments of clarity and growth.\",
                        \"evaluation_value\": \"Neutral Ending\",
                        \"images\": [
                            \"campusEntrance\",
                        ],
                    }
                ]
            }
        ```
    - Bad Ending:
        ```json
            {
                \"scene\": [
                    {
                        \"scene_num\": \"Scene 20\",
                        \"scene_type\": \"End\",
```

```
                                \"scene_text\": \"The end of the academic year arrives, but instead of relief, you’re left feeling drained and disconnected. The choices you made-isolating yourself and not seeking help-have taken their toll. You realize now that things could have been different, but the weight of depression has been overwhelming, making it hard to break out of the cycle.\",
                                \"scene_inner_monologue\": \"I thought I could handle things on my own, but that only made things worse. Now, as I look back, I see all the moments I could’ve reached out for support, but I didn’t. It’s hard to see a way forward right now, but maybe I need to start doing things differently.\",
                                \"evaluation_text\": \"The consequences of avoiding support and isolating yourself have led to a difficult place. While this ending reflects the real challenges of depression, it’s also a reminder that there’s always an opportunity to change. Acknowledging where things went wrong is the first step toward finding a new path. Even though you feel lost now, recovery is still possible if you choose to reach out and seek help.\",
                                \"evaluation_value\": \"Bad Ending\",
                                \"images\": [
                                    \"campusEntrance\",
                                ],
                            }
                        ]
                    }
                ```
    "
}
```

Prompt 12: User Prompt Few Shot Prompting
```

N Prototype Study - Images



(a) College Tales Logo Image for the Prototype (Generated by DALL-E)

(b) Dorm Image for Scenes on the Prototype (Generated by DALL-E)

(c) Campus Entrance Image for Scenes on the Prototype (Generated by DALL-E)

(d) Canteen Image for Scenes on the Prototype (Generated by DALL-E)

Figure 24: Prototype Static Assets (Generated by DALL-E)

(a) Classroom Image for Scenes on the Prototype (Generated by DALL-E)

(b) College Corridors Image for Scenes on the Prototype (Generated by DALL-E)

(c) Counsellor Room Image for Scenes on the Prototype (Generated by DALL-E)

(d) Gym Image for Scenes on the Prototype (Generated by DALL-E)

Figure 25: Prototype Static Assets (Generated by DALL-E)

(a) Library Image for Scenes on the Prototype (Generated by DALL-E)

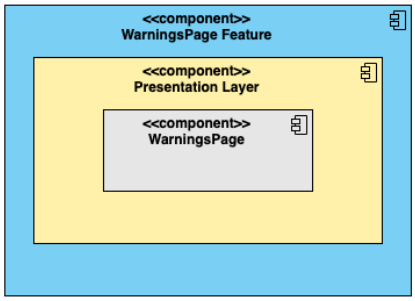
(b) Park Image for Scenes on the Prototype (Generated by DALL-E)

(c) Party Image for Scenes on the Prototype (Generated by DALL-E)

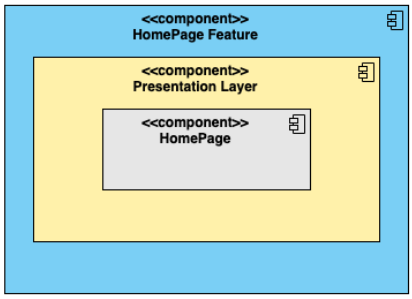
(d) Sports Field Image for Scenes on the Prototype (Generated by DALL-E)

Figure 26: Prototype Static Assets (Generated by DALL-E)

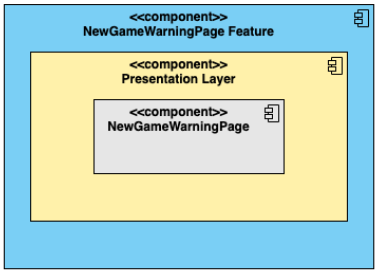
O Prototype Study - Component Diagrams



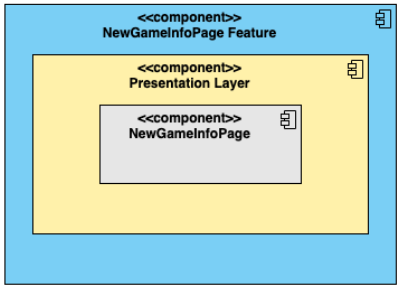
(a) Components Diagram Warnings Page



(b) Components Diagram Home Page

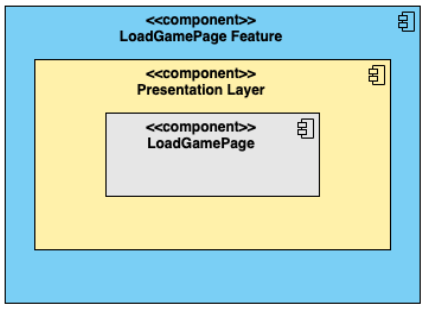


(c) Components Diagram New Game Warning Page

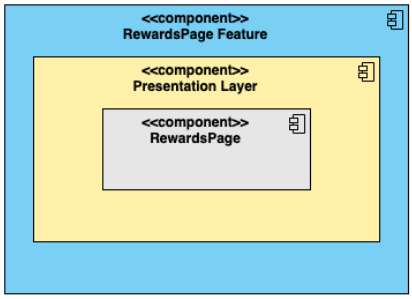


(d) Components Diagram New Game Info Page

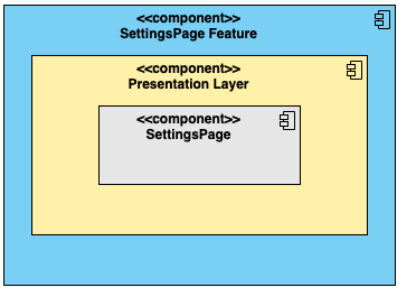
Figure 27: Component Diagrams



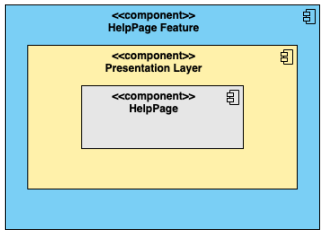
(a) Components Diagram Load Game Page



(b) Components Diagram Rewards Page



(c) Components Diagram Settings Page



(d) Components Diagram Help Page

Figure 28: Component Diagrams