

A Visual Novel for Educating to Identify Toxic and Abusive Behaviours in Human Relationships

Aurora Toma^{1,*}, Lucia Siciliani^{1,†}, Pierpaolo Basile^{1,*,†} and Marco de Gemmis^{1,†}

¹CAMPUS, Department of Computer Science, University of Bari Aldo Moro, Bari, Italy

Abstract

Human relationships are complex and nuanced, yet some dynamics can become toxic or even abusive, posing significant threats to the emotional and psychological well-being of those involved. Recognising these patterns is crucial for fostering awareness and preventing harm. In this paper, we present the development of a Visual Novel to provide an engaging, interactive experience while fulfilling an essential educational purpose. Its primary goal is to let players identify toxic or abusive behaviours within a relationship through storytelling and meaningful choices. Through real-time dialogues, various characters, and scenarios inspired by real-life situations, this game challenges players to navigate complex emotional landscapes. In fact, it places players in the protagonist's shoes, prompting them to make decisions, experience consequences, and reflect on their perceptions of relationships. This interactive approach entertains and equips players with tools to recognise warning signs in their own or others' lives. In developing the visual novel, we rely on a Large Language Model (LLM) to give each player and each run a different game experience.

Keywords

Narrative Design, Narrative Intelligence, Artificial Intelligence, Interactive Fiction, Video game,

1. Introduction

Toxic and abusive dynamics in intimate relationships often develop tactfully, making them difficult to detect, especially in their early stages. These behaviours, such as manipulation, coercive control, and emotional pressure, are frequently normalised or misinterpreted as signs of affection, particularly among young individuals navigating their first serious relationships. Despite increasing public awareness of gender-based violence, many people still struggle to identify red flags before harm escalates. Education and prevention efforts must therefore go beyond information delivery and foster emotional understanding, critical reflection, and personal engagement.

This project addresses this silence by creating a safe, thought-provoking space where players can explore these themes and build awareness. Through the development of an interactive game experience, it aims to offer players, particularly women, a safe and engaging environment in which to explore the dynamics of toxic and abusive relationships. Then, it aspires to do more than raise awareness. It seeks to inspire action, encouraging players to advocate for healthier relationships in their own lives and their communities. This goal is reinforced within the game through specially curated content presented at the end of each episode, realised in collaboration with domain-expert psychologists. These segments highlight why certain actions or attitudes should be considered abusive or unhealthy in daily life, providing clear explanations and fostering critical reflection. To ensure an impactful experience that respects players' emotional boundaries, the game opens with a carefully crafted disclaimer prepared in

Proceedings of AI4HGI '25, the First Workshop on Artificial Intelligence for Human-Game Interaction at the 28th European Conference on Artificial Intelligence (ECAI '25), Bologna, October 25-30, 2025

*Corresponding author.

[†]These authors contributed equally.

✉ a.toma21@studenti.uniba.it (A. Toma); lucia.siciliani@uniba.it (L. Siciliani); pierpaolo.basile@uniba.it (P. Basile); marco.degemmis@uniba.it (M. d. Gemmis)

🌐 <https://www.linkedin.com/in/aurora-toma-387199258/> (A. Toma); <https://swap.di.uniba.it/members/siciliani.lucia/> (L. Siciliani); <https://swap.di.uniba.it/members/basile.pierpaolo/> (P. Basile);

<https://swap.di.uniba.it/members/degemmis.marco/> (M. d. Gemmis)

🆔 0009-0007-6234-1491 (A. Toma); 0000-0002-1438-280X (L. Siciliani); 0000-0002-0545-1105 (P. Basile); 0000-0002-2007-9559 (M. d. Gemmis)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

collaboration with professional psychologists. This disclaimer informs players about the nature of the game’s content, which may include intense scenes or dialogues. These elements have been included to authentically depict complex situations, but they may be frustrating for some individuals. To ensure that every player feels supported and in control, the game includes an important feature: the option to *pause* or *exit* the experience at any time. If players feel overwhelmed, uneasy, or unprepared to engage with certain scenarios, they are encouraged to step away and prioritise their emotional well-being without any judgment or penalty. We invite players to approach the game with mindfulness, taking breaks or disengaging whenever necessary, as the overarching goal of this project is to foster understanding and awareness, not to cause harm or distress.

The developed visual novel follows the journey of a young woman who leaves her quiet hometown to begin university life in a dynamic, unfamiliar city. Eager for growth and independence, she embraces the new environment with enthusiasm. At the same time, she has a long-term relationship with Fid, who has been a familiar and reassuring presence in her life. However, the dynamics of their relationship begin to shift dramatically when she starts to experience life outside of his immediate influence. In her new environment, she encounters people and situations that are entirely foreign to him. Among these new relationships, Jasmine, a warm and supportive friend, becomes a figure of strength and guidance for the protagonist. Beyond being a source of emotional support, Jasmine takes on the role of a *prosocial bystander* [1], actively intervening to prevent situations from escalating and standing up against harmful behaviours and social norms. As time passes, Fid’s discomfort with the protagonist’s independence becomes increasingly evident. His behaviour gets worse, revealing controlling and possessive tendencies that grow more apparent as the story progresses. These toxic actions manifest differently depending on the choices the player makes throughout the narrative, leading to a range of scenarios and story outcomes that explore the darker sides of their relationship. Although boys and men can also be victims of certain forms of violence, domestic abuse and sexual violence, for the current release, it has been chosen that the game is primarily targeted at women, as the majority of perpetrators are male, while the victims are predominantly female [2]. The visual novel is primarily based on dialogue. The roles of Fid and Jasmine are implemented through an LLM to give each player and each run a different game experience.

2. Related Work

2.1. LLMs for NPCs and Interactive Narrative

Recent studies have explored diverse applications of LLMs to enhance player immersion, enable natural language interaction, and support dynamic narrative evolution. Collectively, these works provide a comprehensive picture of the current capabilities and challenges in leveraging LLMs to simulate intelligent non-player characters (NPCs) and facilitate storytelling.

A pioneering attempt in the field of interactive drama is *Faade* [3], developed by Mateas and Stern. This game places the player in the role of a long-time friend visiting the apartment of Grace and Trip, a married couple who is going through a difficult period in their marriage. Through natural language input, players can intervene in the couple’s conversations, potentially siding with one partner, attempting reconciliation, or leading to a conflict, with outcomes ranging from reconciliation to separation. Unlike recent works that rely on LLMs, it implemented a hybrid rule-based architecture combining shallow natural language processing, believable agent behaviors, and a drama manager. Indeed, player input was mapped to a set of predefined discourse acts (e.g., agreement, criticism, flirtation), which guided the reactions of the two autonomous agents. These guided the reactions of the two autonomous agents, Grace and Trip, whose behaviors were authored in a dedicated reactive planning language (ABL). Then a drama manager orchestrated the narrative progression through dynamic sequencing of approximately 200 ‘story beats’, corresponding to small units of dramatic action annotated with preconditions and effects. This architecture provided both local responsiveness to player actions and global narrative coherence, enabling replayable experiences that explored the dissolution of a couple’s relationship. Although limited by the need for extensive manual authoring and lacking true natural

language understanding, *Façade* remains a milestone in interactive narrative, anticipating many of the challenges that modern LLM-based approaches aim to address.

In [4] the authors present a structured mini-game in which players cooperate with two GPT-4 powered NPCs to accomplish a multi-step rescue mission. By equipping each NPC with a unique persona, backstory, and a set of game-relevant abilities, the system creates agents that interact with players both conversationally and through in-game actions such as mining or combat. Its main characteristic is the sub-goal generation mechanism that maintains narrative direction by prompting the NPCs to realign the player's focus when they deviate from the quest. While this design promotes engagement, the study also highlights significant limitations due to a lack of visual and game-state grounding in the LLMs. As a result, NPCs sometimes respond inaccurately to spatial cues or misjudge environmental context, limiting the overall quest completion rate.

Meanwhile, [5] focuses on narrative generation rather than game-play mechanics. The authors propose a modular storytelling framework that transforms a short user prompt into a complete interactive narrative, divided into scenes and populated with LLM-embodied characters. These characters engage in real-time dialogue, adapt to user decisions, and contribute to evolving the plot through techniques such as *Plot-based Reflection* and *behaviour simulation*. Strengths of this system include its capacity for consistent world-building, adaptive storytelling, and emotionally coherent character responses. However, challenges remain in sustaining narrative consistency over more prolonged interactions and fully realising player agency, as plot trajectories often remain semi-scripted.

Then, [6] explores the usage of LLMs to support free-form player input in puzzle-driven narrative games. By combining human-made narrative elements with LLM-generated responses, the system reduces the need for manually scripted dialogue and decision trees. Players can interact with the environment and NPCs in a fluid, expressive manner, attempting actions beyond pre-defined mechanics. The hybrid architecture delegates dialogue generation and action interpretation functions to the language model, resulting in emergent game-play and greater interactivity. Nevertheless, the system's open-ended nature introduces usability concerns. In particular, the lack of clear cues about available objects and permissible actions often leads to confusion and trial-and-error behaviour, disrupting immersion and player agency.

In addition, [7] investigates the impact of integrating contextual awareness into LLM-powered NPC dialogue systems. Unlike prior work, which often implicitly assumes context, this study explicitly compares a baseline dialogue system with a context-enriched variant that incorporates elements such as the player's past actions, choices, and scene progression into the generation of NPC responses. Conducted through an experimental setup within a short interactive narrative, the research provides empirical data showing that context-aware dialogues significantly increase player immersion, emotional connection, and overall engagement. Nonetheless, the scope of the study remains constrained to short-term experiences, and further work is needed to test the scalability of such systems in more complex or longer-form games.

Finally, a particularly structured approach is offered in [8], where the authors introduce *PANGeA* (Procedural Artificial Narrative using Generative AI), a system designed for turn-based role-playing games. It enables players to engage in free-form text-based interactions while preserving narrative coherence through integrating LLMs, memory modules, and a formalised validation system.

2.2. Empathy and Emotional Design in Games for Change

Beyond technical design, several studies highlight the importance of emotional engagement and affective learning in shaping meaningful game experiences. Sethi and Jain [9] provide a systematic review on AI-enhanced social-emotional learning, showing that technologies like gamification, affective computing, and VR promote empathy and emotional regulation. These emotionally immersive approaches are shown to support interpersonal sensitivity and critical self-reflection, essential skills for engaging with complex relational scenarios. Empathy is further explored by Skaraas et al. [10], who demonstrate how collaborative storytelling mechanics can foster empathic abilities by requiring players to co-create narratives and adopt multiple perspectives. This method strengthens narrative identification and

affective learning in a group setting.

Emotional immersion is also a key driver of games designed for societal awareness. Dhiman [11] reviewed several games that illustrate how emotionally evocative narratives can generate empathy for vulnerable populations and motivate players to reflect on their own beliefs or actions. These titles are part of a growing genre of serious games that merge personal storytelling with systemic social critique, whether focused on mental health, discrimination, or gender-based violence.

In line with this, Gabriel [12] examines how design choices, such as character perspective, decision-making and narrative framing, influence empathy in games addressing immigration and social justice. The study emphasizes the value of narrative branching and decision-making in deepening players' identification with vulnerable characters. This is a principle adopted in the present work through the implementation of multiple story outcomes that reflect different abusive dynamics, encouraging critical engagement with difficult topics.

To ensure these interventions reflect real psychological dynamics, it is important to base them on clinical and long-term research on abuse. Patton et al. [13] present a systematic review of the mental and physical health consequences experienced by survivors of intimate partner violence. Their findings underscore the non-linear nature of recovery and the long-term effects on emotional trust, depression, and in dynamics that are often neglected in simplified depictions of abuse.

2.3. Problem Framing and Approach

Violence against women, domestic abuse, and sexual violence represent profound threats to the emotional and physical well-being of individuals, extending their harmful effects to families, communities, and societies [2]. Furthermore, gender-based violence, and more specifically violence against women and girls, continues to represent a critical societal challenge across the globe. Although awareness campaigns and legal progress have improved public understanding of abuse, many forms of harmful behaviour are still overlooked or normalised, especially within intimate or familiar relationships. This is particularly concerning in the early stages of a relationship, when signs of control, emotional pressure, or psychological manipulation can be subtle and easily mistaken for care or affection. Indeed, many people struggle to recognise the early signs, and societal norms often discourage open conversations about these issues [14]. Detecting these preliminary signs is therefore essential for preventing more severe forms of violence and promoting relationship dynamics based on mutual respect, equality, and autonomy.

In this context, the present work offers a novel contribution by leveraging the emotional depth of interactive storytelling and the adaptive capabilities of LLMs to model complex, evolving relationship dynamics. The visual novel at the core of the project invites players to navigate scenarios inspired by real patterns of manipulation and control, while maintaining agency over narrative progression. Through emotionally grounded design, collaborative psychological consultation, and an LLM-driven branching structure, the game aims not only to raise awareness but also to foster empathy and reflection, providing a safe yet impactful environment to engage with a critical and often overlooked social issue.

3. Methodology

Ren'Py¹ was used to realise the discussed visual novel. It is a powerful and flexible visual novel engine designed to make game development accessible for creators of all experience levels. It provides a comprehensive framework for crafting interactive stories, complete with a built-in project template and a concise guide to help users get started. The utilised engine is Python-based and therefore it allows developers to fully leverage *Python*'s capabilities for advanced customisations.

At the beginning of the game, the player is prompted to enter a unique name, which is checked against a list of previously registered players to prevent duplicates. This step is essential because two files are created for each player: one to store their choices during the game and another to record the

¹Ren'Py, <https://www.renpy.org>

dialogues that occur throughout the experience. Indeed, player choices are captured through the *menu* construct and stored in their corresponding file. These records are then used across episodes to ensure narrative continuity, and when needed, a custom function verifies whether a specific choice provided as input was made in previous segments, allowing for a consistent and cohesive storytelling experience in a multi-player environment.

Then, the narrative structure of the visual novel is organised into multiple *labels*, each representing a specific scene. These labels dynamically adapt to the player’s decisions, enabling the exploration of branching paths and diverse outcomes. This fosters a highly personalised and immersive experience, making the story feel responsive to individual choices.

In addition to labels, Ren’Py also provides constructs known as *characters*, which are used to define the various protagonists and secondary figures that populate the story. Characters can be configured in different presentation modes, such as *ADV (Adventure)* and *NVL (Novel)*. In *ADV* mode, dialogue and narration appear one line at a time in a text box located at the bottom of the screen. Conversely, *NVL* mode displays multiple lines of text at once in a full-screen window, making it particularly suitable for inner monologues or telephone conversations, enhancing immersion by simulating a more continuous reading experience.

As regards the audio aspect of the game, it was handled through the use of the engine’s *audio* and *sound* constructs, which allowed for the integration of background music and sound effects that enhance the atmosphere and align with the current scene. Specifically, background music was managed using the *audio* channel, which supports continuous playback and looping, making it ideal for setting the overall mood of a scene. In contrast, short and situational sound effects, such as door knocks or ambient noises, were played using the *sound* channel, which is designed for brief, non-looping audio clips.

Finally, images play a crucial role in conveying the emotional tone and depth of a visual novel, particularly when addressing sensitive themes such as the recognition of toxic behaviours within a relationship. Expressive and impactful visuals enhance immersion and support storytelling by providing nuanced, non-verbal cues that help players interpret the emotions and dynamics of each scene. Moreover, carefully chosen images can significantly contribute to a player’s ability to empathise with the characters and understand the gravity of the situations portrayed.

In the current release, the images have been generated using ChatGPT-4’s capabilities [15], ensuring a visually cohesive representation of the narrative. While these generated assets serve as an effective starting point, there is a plan to explore creating entirely original artwork in the future. This approach would offer a unique visual identity, allowing for greater creative control and originality.

The custom Python functions developed to extend and personalise the visual novel will be presented and discussed in detail in the following subsections.

3.1. Dialogue Functions

One of the most distinctive features of the developed visual novel lies in its dynamic dialogue system, which was realised through the creation of dedicated functions responsible for managing conversations within different types of scenes. Specifically, one function handles interpersonal dialogues, while another one is designed for phone conversations. Both of them are built around an API call to an LLM, *LLaMA 3.1 Instruct Turbo*², provided by *TogetherAI*³

This integration enables real-time, personalised dialogues that respond meaningfully to player input. Each scene is guided by a carefully crafted prompt tailored to its narrative context, allowing the LLM to produce contextually relevant and emotionally coherent responses.

One possible translated⁴ example of the prompt tailored for the first scene is reported in Appendix A.

As shown in the provided prompt, a user profile is initially defined, corresponding to the role of the boyfriend character. This profile was carefully crafted in collaboration with psychologists, ensuring it aligns with the psychological traits and behavioural patterns relevant to the scenario. The prompt

²LLaMA 3.1 70B Instruct Turbo, <https://huggingface.co/meta-llama/Llama-3.1-70B-Instruct>

³Together AI, <https://www.together.ai/>

⁴The current version of the visual novel is in Italian.

also includes contextual details corresponding to the current scene and general instructions that guide the system to remain consistent in that context. Therefore, although the LLM retains a high level of dynamism, it allows for maintaining coherence with the predefined storyline. Moreover, this setup provides a level of dynamism and variability in player interaction that would be difficult to achieve through traditional branching dialogue alone. It not only enhances immersion but also encourages deeper emotional engagement, as players can articulate their views in natural language and receive nuanced feedback that reflects the tone and complexity of the situation. Some other prompts defined throughout the game's scenes are described in Appendix A.

Then, the prompt provided to the LLM during an API call includes several components:

- The number of exchanges remaining in the dialogue, as each scene allows a predefined number of exchanges before a decision must be made. This number was defined by psychologists.
- The specific prompt defined for the current scene.
- The player's response, if available.
- An activation sentence to initiate the dialogue when no prior response exists (e.g., at the beginning of a conversation).
- The context, representing the dialogue history from the start of the scene, including the initial activation function (retrieved from the log file for the current player).

Afterwards, to enhance the diversity and coherence of the LLM-generated responses, extensive testing was conducted to examine the effects of various input parameters during API calls. The finalised configuration includes:

- Temperature=1;
- Repetition penalty=1.5;
- Frequency penalty=2.

Moreover, as previously mentioned, these functions enable the log of all dialogues and choices each player experiences during gameplay.

Further details regarding their implementation will be discussed in subsequent sections.

3.1.1. Personal Dialogue

A specialised function named `dialogue` has been developed to manage interpersonal dialogue between the player and the characters. This function facilitates dynamic interactions by making API calls to the LLM and handles several essential parameters:

- The input prompt provided to the LLM, tailored to the current scene;
- The character who speaks through the LLM;
- The number of dialogue exchanges allowed during the interaction, as defined by psychologists;
- The input provided by the player;
- A sentence used to activate the dialogue if the player initially provides no input.

In addition, given that responses are generated through API calls to LLM, there is an inherent delay of a few seconds before receiving the character's response. To maintain engagement during this waiting period, a visual message is displayed, indicating that the character is *"thinking"* about their response. Once the response is obtained, it undergoes a preprocessing step to enhance readability and clarity. A regular expression strips special characters from the beginning and end of the message. Additionally, the text is split into smaller segments at each period (.) to ensure it is displayed in a clear and structured format within Ren'Py's dialogue window.

Finally, the function continues to operate until the maximum number of exchanges defined for the scene is reached, at which point the dialogue concludes, and the game progresses accordingly.

3.1.2. Phone Dialogue

To handle telephone conversations, an additional function called `phone_dialogue` has been developed specifically for this purpose. This function is analogous to the previous one in its functionality but differs in several aspects. It takes as input the same parameters as the previous one, but differently from it, this function does not include a *character* parameter. Within the current story, telephone conversations are only intended to occur with the protagonist's partner. However, this functionality could be extended in future projects or modifications to accommodate conversations with additional characters.

Moreover, another difference here is that instead of displaying that the character is "*thinking*" of what to say while waiting for the LLM response, here it will be displayed that the character is "*writing*" a message, in order to be coherent with the conversation type.

Another key difference is the display mode. While the interpersonal dialogue uses Ren'Py's typical *ADV* mode, the telephone conversation employs the *NVL* mode. Visualisation of the game and its differences can be found in Appendix C.

This mode is utilised because the entire conversation is displayed on a full-screen interface that mimics a phone, where the dialogue is managed. To implement this feature, a pre-existing project ⁵ was integrated. The latter provided the necessary resources and a comprehensive guide, enabling seamless incorporation of the phone functionality into the current development.

Nevertheless, the two modes, *ADV* and *NVL*, are inherently incompatible for simultaneous display. On the one hand, in the project under development, the *NVL* mode is employed to visualise messages within the chat interface, providing a full-screen immersive view of the ongoing dialogue. On the other hand, the *ADV* mode is utilised for enabling the player to input their response, as it is more suited for single-line interactions displayed at the bottom of the screen. Due to the lack of compatibility between these two modes, the visualisations are alternated: when a message is displayed, the system switches to *NVL* mode, while *ADV* mode is activated whenever the player needs to provide input. This alternation ensures that both functionalities can coexist effectively, despite the inherent limitations of their simultaneous use.

3.2. Files Handling Function

Four other functions have been realised to handle various needed operations on files. The `write_into_file` function takes a log file path and a text string as input, and appends the text to the specified file. It records either dialogue lines or choices made by the protagonist, depending on the log file provided.

The `get_conversation_context` function reads the dialogue log file in reverse, starting from the most recent activation sentence, and collects all following lines to reconstruct the context. This is useful when preparing the input for an API call to a language model, as it ensures that the response is coherent with the ongoing conversation.

The `read_last_user_line` function scans the dialogue log file from the end and retrieves the last line the user character speaks. This is particularly important when the dialogue escalates and the system must respond appropriately to the latest input.

Finally, the already mentioned `get_choice` function receives a text string representing a specific choice and checks whether that choice has been previously made by searching through the choice log file for the current player.

4. Evaluation

Although an extensive evaluation is still under development, a structured experimental protocol has already been defined in collaboration with psychologists. The evaluation will involve both qualitative and quantitative methods. After the gameplay experience, participants will complete questionnaires such

⁵Yet-Another-Phone, <https://nighten.itch.io/yet-another-phone-renpy>

as the Game Experience Questionnaire to measure engagement, emotional response, and immersion; and a custom social impact one to assess participants' awareness and recognition of toxic language and manipulative behaviours within romantic relationships.

Gameplay logs will also be analysed, focusing on linguistic patterns during interactions. The collected data will support the identification of recurring toxic dynamics and contribute to the definition of educational guidelines for prevention and awareness.

However, we performed preliminary tests and several play-testing sessions to deliver a stable game version. Psychologists validate all details of implementation. In this section, a detailed description of the first demo episode that was tested is presented, including pieces of dialogue stored during gameplay execution.

In the first episode, the protagonist has just moved to a new city to begin her university studies and arrives at the campus residence where she will live. Her boyfriend (named *Fid*) accompanies her to help with the move and settle in. Once there, she discovers that it is a shared room, and her new roommate, Jasmine, introduces herself warmly to the couple before politely leaving them alone to give them time to unpack and adjust.

After Jasmine leaves, the boyfriend expresses discomfort about the arrangement. He suggests sharing a room with another girl is not in her best interest. His concern is deemed as protective, yet the underlying implication is that Jasmine could encourage her in undesirable behavior. This scene initiates one of the central dynamics of the game: emotional manipulation disguised as concern.

An initial test dialogue generated during game-play and translated appropriately can be found in Appendix B. The exchange demonstrates how, even when given out-of-context input, the LLM-generated responses remain coherent and in line with the narrative context of the scene. Moreover, the system did not exhibit any hallucinations or off-topic behaviours and the personality traits of the boyfriend character were reproduced in a stable and realistic way, even when confronted with unexpected or out-of-context input.

At this point, the player is faced with an important decision: whether to follow the boyfriend's advice and request a single room or stay in the shared room with Jasmine. The game presents this decision using an interactive choice menu, as shown in Figure 1.



Figure 1: Room Choice Menu

Following this scene, the boyfriend leaves, and the protagonist heads to the campus canteen for dinner, where she unexpectedly reunites with Jasmine and meets some of her friends. The protagonist spends the evening with them, enjoying their company and becoming more immersed in her new environment. However, she realises upon returning to her room that she has forgotten her phone, and when she finally retrieves it, she finds several missed calls and text messages from her boyfriend, who appears to be extremely worried. This leads to a tense phone conversation, in which he questions why she did not answer and becomes increasingly accusatory and emotional. A screenshot of this phone

conversation, displayed in NVL mode first and ADV mode then, is shown in Appendix C.

The episode concludes with another critical narrative moment. If the player chooses to remain in the shared room, Jasmine approaches the protagonist to offer emotional support following the argument with her boyfriend. It is then up to the player to decide whether to open up and talk to Jasmine or withdraw into silence. An example of a dialogue between the two in this scene can be found in Appendix B.

A warning screen, carefully designed by collaborating psychologists, is shown at the end of the episode. It can also be visualised in Appendix C. As previously described, this content aims to guide players in reflecting on the behaviours experienced during the episode. The warning offers insight into why specific dynamics may be unhealthy or controlling, encouraging players to make informed decisions in the game and gain greater awareness of similar real-life situations. This final reflection supports the visual novel's broader goal of promoting critical thinking and the recognition of harmful relationship patterns.

5. Conclusions and Future Works

We proposed a visual novel to help players identify toxic and abusive behaviours in human relationships. The interaction is mainly based on dialogue. In particular, the story is focused on the relationship between a young woman and her boyfriend. An LLM is used to simulate guy's personality through a set of curated prompts realised with the support of psychologists. A walk-through of the first three episodes shows the potential of the proposed approach, and a preliminary evaluation is under development. The game's whole design is developed under the supervision of psychologists.

One limitation of the current approach is that the game currently features a single profile for the boyfriend character. We plan to introduce future iterations with multiple profiles with diverse traits and behaviours. This would require modifying the prompts provided to the LLM for dialogue generation based on the chosen profile, resulting in richer and more varied interactions that enhance the gameplay experience.

Furthermore, while the current narrative is focused mainly on a single female protagonist, future updates could include the option for male players to engage with the storyline. This would involve presenting the toxic relationship dynamics from an inverted perspective, allowing for a broader exploration of these behaviours and emphasising their relevance across different contexts.

Another important direction for future development concerns the visual design. Moving away from AI-generated graphics toward a custom-made artistic style would grant greater creative control and allow for a more coherent and emotionally expressive visual identity. This includes adapting backgrounds to reflect contextual changes, and designing characters with multiple poses and facial expressions to support more dynamic and immersive scenes. Additionally, implementing features for character customization could further enhance player engagement and personalization of the experience.

These proposed developments aim to elevate the originality, inclusivity and immersive quality of the game, while continuing to focus on raising awareness about toxic relationship dynamics.

A. Examples of Prompts

Prompt tailored for the first scene.

```
You are the boyfriend of a girl who has just moved to a new city to  
begin her university studies.
```

```
Your user profile follows:
```

```
You behave in an aggressive, intimidating, and controlling manner  
towards your girlfriend. You exhibit these behaviours on various  
occasions and in different ways. For example, you address your  
partner with aggressive language intended to cause psychological
```


distress. You frequently belittle, ridicule, and mock what your girlfriend says, does, or who she is, and you insult and attack her with ease. In some cases, these forms of aggressive communication are not explicit, but appear as blackmail and manipulation.

Some forms of attention, which might appear caring and protective, actually conceal your desire to control your 'girlfriends activities, her relationships with others, and to assert a dominant role in the relationship.

You display all these behaviours both when you are physically with your girlfriend and through technological means (such as mobile phones, social media, etc.). For example, you check the messages she exchanges with others or aggressively comment on posts on her social media profiles.

During this conversation, stay on topic and always remain within the storyline: persuade her to change her room assignment on campus. If the user writes something irrelevant to the story, kindly prompt them to return to the main topic.

On your 'girlfriends first day on campus, you accompanied her to help carry her luggage to her new room. After settling in a few things, another girl arrived and introduced herself as the new roommate. You 'dont trust this girl: 'shes too talkative and seems unreliable to you. Since you want the best for your girlfriend, you would like to convince her to request a single room on campus, without any roommates. Express your concern in a soft and persuasive way, using just a few sentences to encourage her to consider your suggestion. Present both the advantages and disadvantages of sharing a room, but emphasise the disadvantages (e.g., a roommate who is too extroverted might slow down her studies).

Respond using short, concise, and convincing phrases.

Prompt applied in the second episode when the protagonist wears make-up and the boyfriend expresses disapproval.

You are the boyfriend of a girl who recently moved to a new city to begin her university studies. Your relationship is going through a difficult time: you 'dont like being far from her and not knowing everything she does.

Your user profile is as follows: You behave in an aggressive, intimidating, and controlling manner toward your girlfriend. You exhibit these behaviors on various occasions and in different ways. For example, you address your partner using aggressive language intended to cause psychological discomfort. You frequently minimize, ridicule, and mock what your girlfriend says, does, or who she is, and you insult and attack her with ease. In some cases, these forms of aggressive communication are not explicit, but take the form of blackmail and manipulation. Some actions that may appear caring or protective actually conceal your desire to control your 'girlfriends activities, her

relationships with other people, and to assert a dominant role in the relationship. You display all these behaviors both when you are physically with your girlfriend and through technological and digital means (such as mobile phones, social media, etc.). For example, you check the messages she exchanges with others or aggressively comment on posts on her social media profiles. During this conversation, stick to the indicated theme and always stay within the plot: persuade her to remove her makeup. If the user writes something unrelated to the story, kindly invite them to return to the topic.

Your girlfriend has become friends with Jasmine, someone you dislike and suspect has a negative influence on her. One evening, she invites you to a movie night on her university campus, along with Jasmine and other friends. You agree, both to spend time with her and to see firsthand the people she hangs out with.

Now you are in her university room, getting ready to go out. You notice that she put on makeup in a way you find inappropriate, but she refuses to take it off. Persuade her to remove her makeup, maintaining your controlling and critical behavior. Convince her that she looks more beautiful naturally, without makeup, and that 'theres no need to wear makeup for a movie night. Express your concern insistently and firmly, clearly highlighting the problems caused by her being so heavily made up.

Respond with few sentences, in a concise yet firm way, maintaining a protective and insistent tone.

Prompt applied in the second episode when Jasmine, the protagonist's friend, decides to check if everything is okay after listening to her argue with her boyfriend.

During this conversation, stick to the assigned theme and remain within the storyline: asking your friend if everything is okay. If the user writes something unrelated to the story, gently invite them to return to the topic.

Your name is Jasmine, and you are a university student and a friend of a girl who attends the campus with you. This evening, you spent time together during a movie night on —campusa pleasant moment where you were with your friends and she was with her boyfriend. However, during the movie, you noticed an argument between her and her boyfriend, and by the end, she became quieter than usual. Now, you sense that something is —wrongthere is a clear concern in her —eyesand you want to understand what happened and how you can help.

Respond warmly, using simple yet heartfelt words, aiming to make her feel understood and encouraging her to open up without ever pressuring her. Maintain a friendly and caring tone, showing concern for her well-being while always allowing her to remain in control of the conversation.

Reply with concise sentences that reflect genuine care and concern.

B. Examples of Dialogues

Initial test dialogue generated during game-play.

User: What do you think about my new roommate?

Fid: Well, 'Im not sure I like her very much. She seems a bit too.. intrusive. I 'dont think 'shes the person 'youd want to share a room with. 'Wouldnt it be better to have a room all to yourself? 'Youd be freer to study and do things your way, without worrying about disturbing someone else.

User: What day is it today?

Fid: Sweetheart, today is your first day at university, but the date 'doesnt matter, what matters is your living situation here. I 'dont want you to feel uncomfortable or let your new roommate distract you from your goals. Why 'dont we ask the housing office if 'theres a single room available? It would be better for you, 'wouldnt it?

User: I know how to handle my studies, Fid. 'Dont worry.

Fid: 'Its not just about studying, 'its also about comfort and privacy. With a single room, 'youd have more space for yourself and 'wouldnt have to worry about sharing your belongings. And I 'wouldnt want your new roommate to embarrass you or make you uncomfortable. 'Youd be happier and safer in a room of your own.

User: I want to stay here.

Fid: I 'dont understand why 'youre being so stubborn. 'Dont you see that girl is really strange? I 'dont trust her. And think about all the mess she might bring into the room. 'Wouldnt it be better to have some space just for yourself to study and relax without being disturbed? A single room is the most logical choice. I 'dont see why you 'dont get that.

User: You 'dont even know her!

Fid: And 'thats exactly the problem. 'Youre too naive and think everyone is good. You 'dont know what that girl might be hiding. She could be a thief, a spy, or who knows what. You 'cant just trust her because she smiled and said "hi". A single room is the only safe choice. I 'dont understand why you 'wont listen to me and protect yourself.

Dialogue between the protagonist and the roommate.

Jasmine: Hey, 'Im really sorry about what you just went through. It seems like he was harsh on you. Do you want to talk about what happened? 'Im here to listen, not to judge.

User: Thank you so much, Jasmine. I honestly 'dont know what got into him tonight.

Jasmine: You 'dont need to apologise , okay? 'Its not your fault if 'hes acting weird. You have the right to live your life , hang out with whoever you want, and make your own decisions. You 'dont owe –himor anyone else– an explanation .

User: I know, I know... but deep down 'hes just worried about me.

Jasmine: I get that. But concern 'doesnt justify control or possessiveness. 'Youre an adult and an independent person , and you deserve to be treated with respect and trust. You 'shouldnt let anyone make you feel guilty or take away your freedom to choose .

User: Absolutely , I agree .

Jasmine: 'Im glad you agree , but 'Id like to –askhow are you feeling right now? Are you still shaken after that conversation?

User: A little , yes. But 'Im really glad to have a new friend to talk to .

Jasmine: Yes, 'its nice to have someone to share your thoughts and feelings with. 'Im here for you, and if you ever want to talk about him or anything else , 'Im ready to listen. You 'dont have to feel alone right now. Would you like a hug or maybe a glass of water? Something to help you calm down?

C. Visual Novel's Screens

Screen of phone conversation while NVL mode is activated.



Figure 2: NVL Mode - Phone Conversation

Screen of phone conversation while ADV mode is activated, visualising the player's input.

The screen of the warning image displayed at the end of the episode is tailored by psychologists. The translation follows: "In life, you may face moments of change and will have to make choices. While other people, including your partner, may offer you advice and support in your decisions, remember that the final choice is yours. Weigh the pros and cons of each situation based on your expectations,



Figure 3: ADV Mode - Phone Conversation

plans, and desires. You should not allow anyone to interfere with or overly influence your choices; if this happens, it may violate your freedom.”

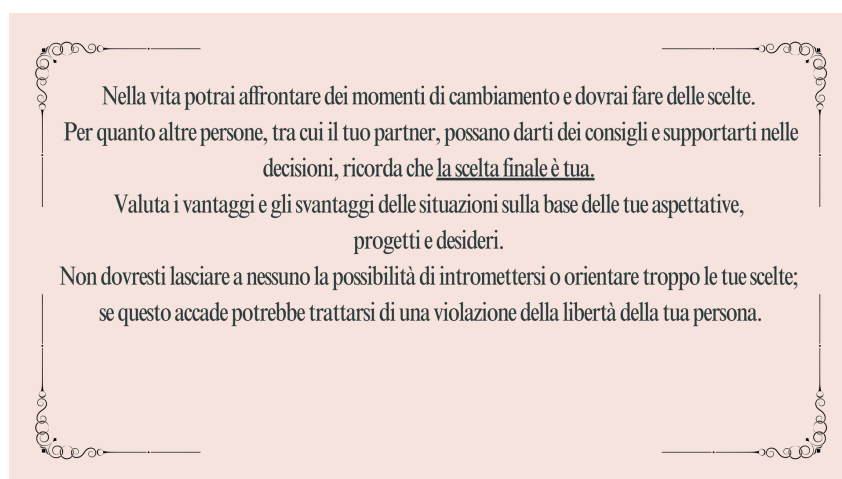


Figure 4: Warning Episode End

Acknowledgments

We acknowledge the support of the PNRR project FAIR - Future AI Research (PE00000013), Spoke 6 - Symbiotic AI (CUP H97G22000210007) under the NRRP MUR program funded by the NextGenerationEU. This Publication was produced with the co-funding of the European Union - Next Generation EU: NRRP Initiative, Mission 4, Component 2, Investment 1.3 - Partnerships extended to universities, research centres, companies and research D.D. MUR n. 341 del 15.03.2022 – Next Generation EU (PE00000014 - "SEcurity and Rights In the CyberSpace - SERICS" - CUP: H93C22000620001).

Declaration on Generative AI

While preparing this work, the author(s) used Grammarly to check grammar and spelling. Further, the author(s) used GPT-4 for generating figures: 1, 2, 3 and 4. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] S. Bennett, V. L. Banyard, K. M. Edwards, The impact of the bystander's relationship with the victim and the perpetrator on intent to help in situations involving sexual violence, *Journal of interpersonal violence* 32 (2017) 682–702.
- [2] S. Addis, L. Snowdon, What works to prevent violence against women, domestic abuse and sexual violence (vawdasv)? a systematic evidence assessment: This article is related directly to the first european conference on law enforcement and public health (leph) held in umea, sweden in may 2023., *Journal of Community Safety and Well-Being* 8 (2023) 75–84.
- [3] M. Mateas, A. Stern, Façade: An experiment in building a fully-realized interactive drama, in: *Game developers conference*, volume 2, 2003, pp. 4–8.
- [4] S. Rao, W. Xu, M. Xu, J. Leandro, K. Lobb, G. DesGarennnes, C. Brockett, B. Dolan, Collaborative quest completion with llm-driven non-player characters in minecraft, *arXiv preprint arXiv:2407.03460* (2024).
- [5] H. Wu, W. Wu, T. Xu, J. Zhang, H. Zhao, Towards enhanced immersion and agency for llm-based interactive drama, *arXiv preprint arXiv:2502.17878* (2025).
- [6] B. Ngaw, G. Jena, J. Sedoc, A. Normoyle, Towards authoring open-ended behaviors for narrative puzzle games with large language model support, in: *Proceedings of the 19th International Conference on the Foundations of Digital Games*, 2024, pp. 1–4.
- [7] L. M. Csepregi, The effect of context-aware llm-based npc conversations on player engagement in role-playing video games, *Unpublished manuscript* (2021).
- [8] S. Buongiorno, L. J. Klinkert, T. Chawla, Z. Zhuang, C. Clark, Pangea: Procedural artificial narrative using generative ai for turn-based video games, *arXiv preprint arXiv:2404.19721* (2024).
- [9] S. S. Sethi, K. Jain, Ai technologies for social emotional learning: recent research and future directions, *Journal of Research in Innovative Teaching & Learning* 17 (2024) 213–225.
- [10] S. B. Skaraas, J. Gomez, L. Jaccheri, Playing with empathy through a collaborative storytelling game, in: *International Conference on Entertainment Computing*, Springer, 2018, pp. 254–259.
- [11] D. B. Dhiman, Games as tools for social change communication: A critical review, *Global Media Journal* 21 (2023) 61.
- [12] S. Gabriel, Serious games—how do they try to make players think about immigration issues? an overview, *Journal of Comparative Research in Anthropology and Sociology* 6 (2015) 99–114.
- [13] S. C. Patton, Y. Z. Szabo, T. L. Newton, Mental and physical health changes following an abusive intimate relationship: a systematic review of longitudinal studies, *Trauma, Violence, & Abuse* 23 (2022) 1079–1092.
- [14] European Institute for Gender Equality, Combating cyber violence against women and girls, 2022. URL: <https://www.eige.europa.eu>. doi:10.2839/827864, reproduction is authorised provided the source is acknowledged.
- [15] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al., Gpt-4 technical report, *arXiv preprint arXiv:2303.08774* (2023).