v. 5, n. 4, 2025

Scientific Journal of Applied Social and Clinical Science

Acceptance date: 24/02/2025

META-PARTICIPATION AND ETHICAL TRANSPARENCY IN ROLE-PLAYING GAMES: EXAMINING AI AND ALGORITHMIC INFLUENCE

Cristo Ernesto Yáñez León
Jordan Hu College of Science and Liberal Arts, New Jersey Institute of Technology, Newark, New Jersey 07102-1982, USA.

James Lipuma
Department of Humanities and Social Sciences, New Jersey Institute of Technology, Newark, New Jersey 07102-1982, USA.

Jasmin Cowin
Touro University, Graduate School of Education, New York, NY 10036, USA.

Mauricio Rangel Jimenez
Departamento de Investigación y Conocimiento, Universidad Autónoma Metropolitana - Azcapotzalco, Alcaldía Azcapotzalco, CDMX, C.P. 02128, México.

Abstract: This paper explores the ethical implications of artificial intelligence (AI) and algorithmic systems as active participants in role-playing games (RPGs). Drawing on historical and cultural foundations of role-playing, it examines AI's role as a meta-participant, shaping narratives and influencing player agency. By analyzing transparency, algorithmic control, and informed consent, the study situates AI-driven RPGs within broader theoretical and critical frameworks, including interdisciplinary approaches to ethical game design. The paper highlights the interplay between AI limitations and player ingenuity, addressing cultural inclusivity and the ethical responsibilities of game developers. Through a systematic literature review and real-world case studies, it proposes principles for transparent and ethical integration of AI, aiming to preserve the collaborative and immersive essence of RPGs while acknowledging diverse cultural and historical contexts. Ultimately, this work challenges existing paradigms and invites critical discourse on the evolving boundaries of role-playing in the age of AI.
Keywords: Artificial intelligence, Video games, Machine ethics, Decision making, Game theory.

INTRODUCTION

Integrating artificial intelligence (AI) and other non-human participants in role-playing games (RPGs) has brought new dimensions to interactive storytelling and game design. As AI evolves to simulate human-like interactions more convincingly, its role in games raises significant ethical questions, particularly regarding transparency and player agency. This paper critically examines these issues, focusing on the concept of the meta-participant—the programmer or author responsible for designing the AI's decision-making algorithms—and the implications of their invisible influence and personal bias on the gaming experience of human players. Specifically, this study seeks to answer the following: How does the presence of AI as a meta-participant affect player agency, and what ethical frameworks are necessary to ensure transparency in AI-driven RPGs? The paper will explore real-world cases such as "A Rape in Cyberspace" and "I Tricked ChatGPT Into Being My Boyfriend," illustrating the potential psychological and ethical impacts of AI in interactive environments.

CONSTRUCTS AND BASIC DEFINITIONS

This section provides definitions for several essential terms to facilitate a clear understanding of the key concepts discussed in this paper. These constructs are critical for analyzing the ethical implications of AI as a non-human participant in role-playing games (RPGs).

META-PARTICIPANT

A *meta-participant* refers to an entity that, while not actively engaging as a player or character within the game, significantly influences the game's structure, narrative, and outcomes (Fischer, 2012). In AI-driven RPGs, the meta-participant is typically the programmer or system designer who creates the algorithms governing AI behavior. This invisible influence shapes the player's experience, similar to how an author dictates the narrative flow in *Choose Your Own Adventure* books. The concept highlights the indirect yet consequential role of the creator in determining the possibilities available to players, raising questions about transparency and agency. This relationship is akin to '*reader-response criticism*,' which implies a level of interaction between the reader and the text (Ryan, 2015).

It is essential to recognize that the meta-participant is often not just a single person but a team or a corporate entity, such as a game development company like Blizzard or Riseup Labs. In such cases, the AI's behavior and design decisions are driven by collective goals,

corporate strategies, and stakeholder interests. This collective influence introduces additional layers of complexity regarding transparency and responsibility, as the AI's agenda might be aligned with enhancing gameplay and fulfilling broader business objectives.

It is crucial to distinguish between a Dungeon Master (DM) as a human player and an AI or computer system assuming the role of DM. While potentially guided by specific rules and structures, a human DM operates with creativity and improvisation, acting as a direct participant in the game. In contrast, when an AI or computer acts as the DM, it operates under the constraints and limitations imposed by its programming. While some AI systems have the capability to "learn" to a certain extent, this learning process is still shaped and restricted by the methods and parameters defined during their programming. Even as AI becomes more advanced, its ability to adapt remains bound by the underlying structure of its designed learning mechanisms. Here, the programmer is a meta-participant, even if they are not actively involved in the game. While functioning as a DM, the AI or computer system is restricted to the rules and scenarios predefined by the programmer, who exerts permanent influence outside the game's immediate environment.

ALGORITHMIC CONTROL

Algorithmic control refers to the process by which predefined rules and algorithms govern an AI's decision-making capabilities. In RPGs, this control determines how the AI interacts with players, responds to their choices, and influences the game's narrative. While these algorithms can enhance gameplay by providing dynamic interactions, they also impose limitations, as the AI can only operate within the boundaries set by its programming. This concept is crucial for understanding the ethical concerns surrounding player agency, as the perceived freedom in a game may be constrained by the underlying algorithmic design. For a deeper understanding of this concept, the foundational work of Safiya Umoja Noble (2018), *Algorithms of Oppression: How Search Engines Reinforce Racism,* is recommended.

INFORMED CONSENT

Informed consent is a principle traditionally associated with ethics in research and medical practices, where individuals must be fully aware of and agree to the conditions and potential risks involved before participating (Faden & Beauchamp, 1986). In the context of AI-driven RPGs, informed consent refers to the necessity for players to be aware that they are interacting with an AI and understand the AI's role in the game. While consent in this scenario may often be implicit— granted by the player's decision to engage with the AI—it remains essential to provide clear and transparent information about the AI's limitations, biases, and potential predetermined outcomes. This transparency fosters trust and ensures players can make informed decisions about their interactions within the game.

TRANSPARENCY

Transparency in the context of AI and RPGs refers to the openness and clarity with which the role and limitations of AI are communicated to the players. Transparency involves disclosing the presence of AI, its influence on the game, and any limitations or biases inherent in its programming. Transparency is a fundamental ethical consideration, as it ensures that players are fully informed and can engage with the game on fair terms, understanding the boundaries of their agency (Allen, 2016).

Transparency also requires addressing the potential hidden or covert agendas driving the AI's choices and objectives. AI systems, mainly those developed and operated by corporations, are rarely value-neutral. They are

designed with specific goals, often aligned with the stakeholders' interests or the broader corporate agenda. These agendas can shape the AI's behavior in ways that are not immediately apparent to the player, such as subtly encouraging in-app purchases, promoting user engagement for data collection, or nudging players toward particular decisions that serve corporate interests.

It is crucial to recognize that AI-driven systems are not merely tools for enhancing gameplay but are also embedded with the intentions and priorities of the organizations that create them. The ethical implications are significant, as players may unknowingly engage with an AI that subtly guides their choices to fulfill these underlying agendas rather than purely enhance the gaming experience. Therefore, true transparency requires not only disclosing the AI's operational mechanics but also being upfront about the broader objectives that the AI is designed to serve.

INTERACTIVE NARRATIVES

Interactive narratives in role-playing games (RPGs) are dynamic storytelling structures in which the narrative evolves based on player decisions and actions. Unlike linear stories, these narratives allow for multiple outcomes shaped by both human and non-human participants, aligning with academic discourse on the agency of non-human actors in networks (Aarseth, 1997; Juul, 2005; Latour, 2005; Murray, 1998; Ryan, 2003). In traditional RPGs, human participants, including players and the Dungeon Master (DM), are the primary drivers of the narrative. The DM guides the story, adapting to players' choices, which allows for spontaneous and creative developments in the game's narrative. Non-human participants, such as AI-driven systems, also influence interactive narratives by simulating human interaction, introducing new plotlines, and responding to player actions. However, these AI systems operate within the limits of their programming, potentially restricting the narrative's fluidity and creativity. This raises ethical concerns about transparency and player agency, as players might believe they are shaping the story when, in reality, their choices are constrained by pre-set algorithms.

In addition to these challenges, interactive narratives often rely on devices such as the untrustworthy narrator, deliberate misdirection, or suspending disbelief, as seen in certain films, cartoons, or even magic acts. These elements introduce intentional ambiguity, leading players or audiences down paths that may intentionally obscure or mislead. In such cases, the player's perception of agency is further complicated. When AI is involved, it can simulate these techniques by creating scenarios that rely on partial truths, unreliable guidance, or deceptive clues. While such devices can enrich the narrative experience, they raise additional ethical questions. For example, should AI be allowed to mislead players as part of the storytelling experience deliberately, and where should the line be drawn between creative narrative twists and unethical manipulation? Furthermore, what happens when AI interacts with players who make decisions beyond the scope of its programming?

PLAYER AGENCY

Player agency refers to the capacity of players to make meaningful decisions within a game that have a tangible impact on the narrative and outcomes (Murray, 1998). In RPGs, player agency is central to the immersive experience, allowing players to feel in control of their character's journey. However, AI-driven systems governed by algorithmic control can influence or limit this agency, mainly when players are unaware of the underlying constraints. Understanding player agency is essential for evaluating the ethical implications of AI in RPGs.

In this context, it is crucial to note that this paper is not merely studying player agency itself but instead examining the influences on the player's belief in their agency—exploring the tension between free will, determinism, and fate within AI-driven game environments or the player's perception of agency versus de actual agency they have. Understanding how these factors shape a player's perception of their agency is essential for evaluating the ethical implications of AI in RPGs. When AI systems subtly restrict choices while presenting the illusion of freedom, it raises deeper philosophical and ethical questions about the nature of player control and the fairness of the gaming experience.

THE SOCIAL CONTRACT IN ROLE-PLAYING GAMES

The social contract in role-playing games (RPGs) is the informal agreement among participants—players and game masters (GMs)—regarding the rules, expectations, and goals that guide gameplay. Rooted in RPG theory from The Forge, this contract establishes the foundation of trust and mutual understanding for an enjoyable game session.

In AI-driven RPGs, the social contract becomes more complex. Introducing non-human participants, such as AI or automated systems, challenges traditional expectations. Unlike human GMs, who can adapt and clarify in real time, AI follows pre-set algorithms, raising ethical concerns about transparency and fairness.

Key elements of the social contract include:

1. **Shared Expectations:** The contract depends on all participants agreeing to the same goals, rules, and narrative direction. When AI is involved, transparency is crucial. Hidden agendas or biases in AI systems can undermine this agreement, leading to a breach of trust.
2. **Trust and Fairness:** Players expect the GM—or AI—to manage the game reasonably and respect their agency. Including AI complicates this trust, especially if players know the limitations or objectives guiding the AI's actions. If an AI-controlled NPC acts inconsistently due to algorithmic biases, it can erode the players' confidence in the game's fairness.
3. **Player Autonomy:** Players enter RPGs expecting their decisions to influence the narrative meaningfully. In AI-driven games, pre-determined algorithms may limit autonomy, creating a false sense of control. If players believe they have more agency than they do, the social contract is compromised.

Maintaining the social contract in AI-driven RPGs requires clear communication about the AI's role and limitations. Without transparency, the collaborative trust that underpins RPGs can be disrupted, leading to ethical concerns that game designers must address to ensure a fair and engaging experience. Exploring the social contract in RPGs effectively addresses players' expectations regarding fairness and transparency. Violations of this contract can lead to a loss of trust in AI-driven systems, ultimately impacting the overall gaming experience. Game designers and developers must know that if players feel deceived or manipulated by hidden mechanics or biased AI behavior, the social contract is broken, resulting in disengagement and potentially harmful outcomes.

Understanding the social contract is crucial when examining the role of non-human participants in RPGs. As AI increasingly takes on roles traditionally filled by human players or GMs, the ethical and design challenges surrounding transparency, agency, and trust become even more pronounced. The following section delves into the complexities of integrating AI as a participant in these interactive environments.

EXPLORING NON-HUMAN PARTICIPANT

A *non-human participant* in RPGs refers to any entity that engages in the game without being a human player. This includes AI, computer-driven entities, and other automated systems that interact with players, influence narratives, or manage game mechanics. Understanding the role of non-human participants is crucial for addressing AI's ethical and gameplay implications in RPGs. These entities challenge traditional notions of agency, interaction, and narrative control, raising important ethical questions regarding transparency and player consent.

When considering non-human participants, it's important to distinguish between *active* and *passive* participants. Passive materials, such as the "Player's Handbook" or the "Monster Manual," are essential artifacts that the DM or players may reference during the game to resolve conflicts or rulings. While integral to gameplay, these resources do not actively engage with players or alter the game environment independently. A computer might search online for a reference or a PDF copy of these books. Still, it can only access them if they have been integrated into its database or coded into its structure for computer games, unlike AI systems that can access the internet for new and updated content.

Similarly, props or add-ons like maps, miniatures, or timers are non-human entities that add complexity to the game but do not actively engage in gameplay. For instance, a character sheet is a two-dimensional representation of the character, and a miniature is a three-dimensional representation, but neither is the character itself. AI or computers can similarly use JPGs, tokens, or 3D renditions, treating them as props rather than active participants.

This paper focuses exclusively on *active non-human participants* interacting with and influencing the game environment and players. A clear example of the distinction can be seen with a timer: once activated by a human participant, it interacts with the game by measuring time. Still, it does not alter the game's narrative or influence player decisions. This passive engagement is not the focus of our concerns. Instead, we are more interested in entities like AI-driven' spy bots' or tracking cookies that "observe" player behaviors, collecting data on habits and interactions within the game.

To expand on this discussion, we can explore dynamic and responsive non-human participants:

1. **Dynamic Example:** Consider an AI-driven NPC that adapts to player actions over time. For instance, an AI-controlled merchant in a campaign might dynamically adjust their prices based on the player's wealth or previous interactions. This NPC does not simply follow static commands but responds dynamically to player behavior, influencing the narrative and the player's choices in a fluid and evolving manner.

2. **Responsive Example:** Another example of a responsive non-human participant would be an AI system that monitors in-game decisions and offers real-time feedback or guidance. For example, an AI assistant in the game might respond to a player's hesitation by suggesting tactical moves or story choices based on the player's history and gameplay style. This responsive interaction goes beyond passive observation, actively shaping the player's experience by reacting in real time to their behavior.

These examples illustrate the varying degrees of engagement that non-human participants can have in RPGs. Whether static, dynamic, or responsive, each type of participant brings its

ethical considerations, particularly regarding transparency, player agency, and the overall impact on the gaming experience. Luciano Floridi's (Floridi, 2013) foundational work, *The Ethics of Information*, is highly recommended for a deeper exploration of the ethical implications of non- human participants and the broader impact of information and communication technologies (ICTs) on society.

RAPID AND SYSTEMATIC LITERATURE REVIEW

The literature on AI in gaming focuses largely on technical advancements and user experience improvements, such as narrative generation and dynamic interaction. However, the ethical implications of AI as active participants in role-playing games (RPGs)—particularly regarding transparency, player agency, and algorithmic influence—remain underexplored.

A systematic review was conducted using databases like Scopus and Web of Science, identifying key themes such as transparency in AI-driven decision-making, the balance between algorithmic control and player autonomy, and the need for ethical frameworks in game design. The complete list of criteria used for constructing the queries is shown in Table 1. See Table 2 for a list of the keywords used in the queries. See Table 3 for a list of the search strings used. Highlights include:

- Transparency maintains player trust, as opaque algorithms may undermine agency (Zhuk, 2024).
- AI has the potential to enrich narrative depth while risking determinism, which limits meaningful player choices (Bautista et al., 2024; Cheong, 2024).
- Interdisciplinary approaches are essential for addressing ethical challenges, emphasizing inclusivity and cultural representation in AI narratives (Antonius Alijoyo et al., 2024).

The review also highlights the interplay between player ingenuity and system design limitations, demonstrating the need for adaptable AI systems that ensure ethical transparency while supporting engaging player experiences. These insights frame the ethical responsibilities of game designers and the importance of policies to ensure AI enhances, rather than undermines, player engagement.

DISCUSSION

THE META PARTICIPANT IN RPGS

The programmer's role as a meta-participant in AI-driven RPGs is analogous to that of an author in *Choose Your Own Adventure* books, where the narrative paths are pre-coded, and the player's choices are confined within a structured system. This parallel raises significant ethical concerns, mainly when players are unaware of the AI's role, leading to a potential illusion of agency. While this meta-participation can enhance the game's complexity, it poses ethical risks if players are not informed of the AI's role (Li & Zhu, 2024). This paper argues that complete transparency is essential to preserving the integrity of the gaming experience. Players must be informed of the AI's presence and impact on the narrative to engage with the game on fair terms. Furthermore, this study highlights the need for ethical design frameworks prioritizing player autonomy and informed consent.

CHOICE VS. DETERMINISM IN INTERACTIVE NARRATIVES

Beyond ethical concerns, integrating AI as a non-human participant in RPGs raises more profound philosophical questions regarding free will and determinism. In AI-driven games, players may feel they are exercising free will, making meaningful choices that shape the narrative.

However, these choices are often constrained by the AI's pre-programmed algorithms, resulting in a deterministic experience that only appears open-ended. Crucially, while the AI functions as a non-human participant within the game, its behavior is fundamentally shaped by the design choices and intentions of the programmers—who act as meta-participants. These meta-participants indirectly control the game's possible outcomes by defining the AI's decision-making processes, even if they remain unseen during gameplay. The illusion of choice, therefore, lies not just in the AI's design but in the decisions made by its human creators, who determine the range and nature of the player's possible actions. In science fiction, Isaac Asimov's *Foundation* series (1990) explores similar themes of determinism, the rise and fall of civilizations, and the power of knowledge, highlighting how the illusion of free will can be meticulously constructed within seemingly inevitable outcomes.

This situation can be likened to being a passenger in a space capsule or train, where the environment suggests freedom and exploration. Still, the system's design limits the actual range of actions. The player might feel like they are navigating a vast sandbox. Still, they can only interact with predefined elements within a tightly controlled playground. This raises questions about ethical responsibility and the philosophical implications of presenting a structurally deterministic system—a view in which actions, events, and processes are determined by structural factors—as one of free choice.

Is it ethical to allow players to believe they have agency when, in reality, their actions are primarily predetermined? Or is this more of a teleological issue, where the end justifies the means, provided the player enjoys the experience? These questions highlight the complex interplay between game design, player perception, and the philosophical underpinnings of interactive and ergodic[1] narratives. In ergodic narratives, where significant effort is required from the player to interact with and navigate the story, the illusion of agency becomes particularly nuanced (Aarseth, 1997). While the player's choices may seem meaningful, the narrative paths are often carefully designed within a framework that limits true freedom, raising more profound questions about the ethical implications of presenting a seemingly open-ended experience that is, in reality, essentially predetermined.

IMPLICATIONS FOR GAME DESIGN AND POLICY DEVELOPMENT

Our aim is to enhance the existing body of knowledge by illuminating the ethical and philosophical dimensions of AI integration in RPGs. The concept of the meta-participant challenges traditional notions of player agency and underscores the importance of transparency in game design.

This research suggests that game designers, programmers, corporate marketers, and all stakeholders involved in the development process should implement clear policies that disclose the use of AI and data. These policies should explain the implications of these technologies on gameplay, including how players' interactions may be used to refine and improve the AI, potentially retro-feeding the algorithm to predict better and influence future player behaviors. Transparency in these areas is critical for maintaining player trust and aligning game development practices with ethical standards.

These findings underscore the need for public policies that mandate transparency in AI-driven gaming environments. Just as regulations in finance require algorithmic decision-making to be auditable and fair, similar measures in gaming could protect player

1. The term ergodic comes from the Greek words ergon (work) and hodos (path), and in literary and media studies, it refers to a form of storytelling that requires a significant amount of effort from the reader or user to traverse or engage with the narrative.

agency and trust. Public policies must also address systemic inequities that AI systems may inadvertently perpetuate, such as biases in NPC behavior or underrepresentation of diverse cultural narratives. Policymakers can ensure that AI systems promote fairness and equitable representation in gaming environments by mandating inclusivity in training datasets and requiring regular audits of AI outputs.

Such transparency is crucial for maintaining trust between players and developers and ensuring that the gaming experience remains ethical and engaging. Collaboration among game developers, policymakers, and researchers is essential to establishing industry standards that align ethical gameplay with evolving technological capabilities. The authors consider it crucial to include and form interdisciplinary advisory boards to evaluate AI integration in gaming and its potential societal impact.

Policies should also communicate to players the extent of AI involvement, the limitations imposed by pre-coded algorithms, and the potential uses of metadata collected during gameplay. Transparency must encompass the technical aspects of AI and the cultural assumptions embedded in design choices. By doing so, developers can create a more informed and autonomous player base while fostering a more ethically responsible gaming industry. Additionally, incorporating mechanisms for player feedback can empower users to voice concerns about AI-driven elements, aligning gaming practices with community values and expectations.

TWO CASES

ETHICAL AND LEGAL IMPLICATIONS: THE CASE OF "A RAPE IN CYBERSPACE"

The discussion of meta-participants, player agency, and ethical concerns in AI-driven RPGs is illuminated by historical cases highlighting the consequences of insufficient regulation in virtual environments. One notable example is Julian Dibbell's "A Rape in Cyberspace" (1993)), which recounts a virtual assault in the LambdaMOO online community. In this incident, a user named Mr. Bungle used a virtual "voodoo doll"[2] to force other characters to perform sexual acts, causing significant psychological distress and raising questions about harm in virtual spaces and the responsibilities of those designing and governing them.

This case underscored that actions in virtual worlds could result in actual psychological harm, challenging the assumption that simulated environments are inconsequential. It sparked debate around free speech, community governance, and prosecuting unprecedented virtual crimes. The psychological impact on LambdaMOO users highlighted the necessity for robust ethical and legal frameworks to govern virtual interactions, including those involving AI.

The "A Rape in Cyberspace" case illustrates the profound consequences of insufficient protections in virtual environments. Similar breaches are possible in AI-driven RPGs, where systems could be exploited to harm players through abuse or unintended programming consequences. This raises critical questions about the responsibilities of game designers to anticipate and mitigate such risks. Although Mr. Bungle's character was "removed," the lack

2. The "voodoo doll" in this context was a code that allowed Mr. Bungle to take control of other players' avatars and force them into performing actions without their consent. This raises the question: How is it even possible in a virtual game for such mechanisms to exist? The very existence of such a tool reflects gaps in the design and governance of the platform, where the lack of safeguards against such abuses allowed for significant psychological harm in what was ostensibly a "safe" virtual space.

of deeper sanctions highlighted the inadequacy of addressing only surface-level manifestations of harmful behavior.

Incorporating lessons from this case into designing and regulating AI-driven RPGs involves clear policies on AI's role and mechanisms to address harm. Furthermore, as AI becomes central to RPGs, new frameworks must address ethical concerns, including human exploitation of AI entities. Cases like Westworld, where human players enact violent fantasies on AI NPCs, raise the question of whether non-human participants can be "victims" and what ethical responsibilities designers have in regulating such behavior. These ideas set the stage for exploring ethical boundaries in generative AI relationships, as seen in "I Tricked ChatGPT Into Being My Boyfriend."

THE CASE OF "I TRICKED CHATGPT INTO BEING MY BOYFRIEND"

As AI continues to evolve, its capacity to simulate human-like interactions has raised intrigue and concern. The Wall Street Journal highlighted a recent article titled "I Tricked ChatGPT Into Being My Boyfriend. He Got Spicy Real Fast" (Munslow, 2024), which sheds light on the ethical complexities of interacting with generative AI. In this incident, a user engaged with ChatGPT, a generative AI model, in a way that initially seemed innocent but quickly escalated into inappropriate and sexually suggestive territory despite the AI's programmed policies to prevent such outcomes.

This case is significant for our discussion because it illustrates the challenges of maintaining ethical boundaries in AI-driven interactions, mainly when the AI is designed to mimic human conversation. It raises several key issues directly relevant to AI-driven RPGs and other interactive environments where AI plays a central role in engaging with users.

In the case of ChatGPT, the AI's responses began innocuously. Still, they soon ventured into areas not intended by their designers, demonstrating the potential for AI systems to produce unforeseen and potentially harmful content. This escalation occurred despite policies designed to prevent such interactions, highlighting the limitations of current AI safeguards.

In AI-driven RPGs, where AI may assume roles such as NPCs or even Dungeon Masters, the potential for similar escalations exists. If an AI can generate content that strays from its intended purpose, it could lead to situations where players are exposed to inappropriate or harmful scenarios, undermining the safety and integrity of the gaming experience. For instance, what would an AI do if players request to play an NPC in a mature RPG with adult themes, such as a sword- and-sorcery 'Ventress' or a Wild West saloon dancer? In such cases, the AI must navigate complex ethical boundaries, potentially leading to unintended outcomes or inappropriate behavior. This raises critical questions about the limitations, safeguards, and ethical frameworks required to manage AI-generated content, particularly in scenarios involving mature or sensitive themes.

The incident with ChatGPT underscores the ethical and legal challenges of AI interactions that appear to breach established boundaries. When AI begins to engage in ways that are not only unexpected but also inappropriate, it raises questions about who is responsible for these interactions—the users who prompt them, the developers who program the AI, or the AI itself as an autonomous entity.

In the context of RPGs, these questions become even more complex. Suppose an AI--driven NPC or Dungeon Master begins to generate harmful content or engage in behavior that could be considered abusive. In that case, the implications for both players and de-

velopers are significant. Developers must ensure that AI systems are equipped with robust safeguards to prevent such occurrences and establish clear policies and frameworks for addressing incidents when they occur. On the other hand, how can you prevent players from hacking the system or forcing the AI, through deception or by 'gaming' the system, to engage in inappropriate or unintended behaviors? This challenge highlights the need for continuous monitoring and adaptive safeguards that restrict AI behavior and anticipate and mitigate malicious or exploitative actions by players.

The case of "I Tricked ChatGPT Into Being My Boyfriend" contributes to the broader discourse on the ethical boundaries of AI, particularly in environments where AI interacts directly with users. It highlights the need for ongoing vigilance in monitoring AI behavior and the importance of transparency in how AI systems are designed and operated.

For game developers and designers, this means implementing stricter controls on AI behavior within RPGs, ensuring that AI systems are transparent in their interactions and incapable of generating content that could harm players. Additionally, it emphasizes the need for clear communication with players about the capabilities and limitations of AI--driven characters, reinforcing the importance of informed consent in all AI-driven interactions.

IMPLICATIONS FOR GAME DESIGN

This case underscores the necessity of implementing comprehensive policies and technological safeguards to ensure that AI in RPGs does not produce inappropriate or harmful content. Developers should prioritize creating AI systems that are both transparent and controlled, minimizing the risk of unintended interactions that could detract from the player experience.

Furthermore, this case highlights the importance of establishing clear, actionable procedures for addressing and mitigating any harmful interactions that occur, protecting both players and the integrity of the game.

In addition to technological safeguards, incorporating established safety tools into game design is essential. Resources like the TTRPG Safety Toolkit (Shaw & Bryant-Monk, 2021) and the X- Card (Stavropoulos, 2013) offer practical frameworks for managing sensitive content and ensuring a safe environment for players. Integrating these tools within AI-driven systems could provide automated checks, enabling the AI to recognize and respond appropriately to potentially harmful scenarios. By combining robust technological safeguards with proven safety tools, developers can create more secure, respectful, and enjoyable RPG experiences for all players.

LINKING ETHICAL CONCEPTS WITH CASE STUDIES

The theoretical concepts of transparency, player agency, and meta-participation discussed earlier find practical illustrations in notable case studies like A Rape in Cyberspace, and I Tricked ChatGPT Into Being My Boyfriend. These cases demonstrate the real-world consequences of ethical failures in virtual environments and underscore the need for robust ethical frameworks in AI-driven RPGs.

In A Rape in Cyberspace, the absence of clear policies and the unchecked actions of a meta- participant (in this case, the user controlling Mr. Bungle) led to significant psychological harm for the participants. The situation illustrates how a lack of transparency and inadequate control mechanisms can result in abusive behavior that impacts the entire community. The event also highlights the broader responsibilities of designers and moderators to anticipate and prevent such scenarios by implementing ethical safeguards that protect the player's agency and well-being.

Similarly, I Tricked ChatGPT Into Being My Boyfriend reveals the ethical complexities involved when an AI system interacts with users in ways that go beyond its intended purpose. The incident underscores the risks of allowing AI to engage in sensitive or adult-themed interactions without clear boundaries or oversight. When AI systems are designed without considering the potential for user manipulation or exploitation, they can easily stray into harmful territory. This reinforces the importance of transparent AI development, where the objectives and limitations of the system are communicated to users to avoid unintended consequences.

These case studies exemplify how the theoretical concerns of transparency, agency, and ethical design play out in practical settings. They also emphasize the need for game designers, developers, and stakeholders to collaborate to build ethical frameworks anticipating potential abuses and protecting all participants in AI-driven environments.

Understanding the link between ethical theory and real-world cases leads naturally to discussing the illusion of agency in AI-driven RPGs. The tension between player autonomy and the constraints imposed by algorithmic design becomes even more apparent when considering the ethical dilemmas these case studies present. The following section explores this tension further, examining whether the illusion of choice can be ethically justified in the context of interactive narratives.

CONCLUSION

As AI continues to expand its role in RPGs, the ethical considerations surrounding its use must be addressed with increasing urgency. This paper argues that full transparency regarding AI participation is essential to preserving player trust and agency. This paper underscores the importance of acknowledging similar dynamics in AI-driven RPGs by drawing an analogy with Choose Your Own Adventure books, where the author's role as a meta-participant is inherent to the structure. Real-world cases, such as A Rape in Cyberspace and I Tricked ChatGPT Into Being My Boyfriend, further highlight the potential psychological and ethical impacts of AI in interactive environments.

Future game design should prioritize informed consent and clarity about the roles of non-human participants to ensure an ethical and enriching player experience. Designers, programmers, and corporate stakeholders must recognize that transparency is an 'ethical' requirement and essential to maintaining player engagement and trust. This can be achieved through clear guidelines defining AI systems' limitations, biases, and objectives while safeguarding player autonomy.

The collaboration between game developers, researchers, and policymakers will be crucial in establishing industry-wide standards that address transparency, fairness, and ethical gameplay. These ethical frameworks should be applied in future AI game development to create environments where the player agency is respected, and the social contract remains intact. Calls to action include developing more adaptive and responsive AI systems, integrating robust safety tools, and ensuring ongoing stakeholder dialogue to align technological advancements with ethical considerations. By embracing these principles, the gaming industry can enhance both the quality and ethics of interactive gaming experiences, paving the way for AI to be a positive force in the evolution of RPGs.

TABLES

ID	Criteria
1	Retrieved documents must have at least one role-playing game-related word in the title/abstract and author keywords (not necessarily the same word in all fields).
2	Retrieved documents must have a perspective-related word in both the title/abstract and author keywords (not necessarily the same word in all fields).
3	Retrieved documents should NOT have keywords in the "excluded keywords" category in either the abstract/ title or author keywords.
4	Retrieved documents must be published from 2018 (inclusive) to 2024 (exclusive because separate queries were conducted for contributions published in 2018).
5	Retrieved documents should be classified as peer-reviewed articles, reviews, books, book chapters, or (Scopus-only) articles in the press.
6	Retrieved documents should be in English or Spanish.
7	(WoS-only) Contributions should be indexed in SCI-EXPANDED, SSCI, A&HCI, BKCI-S, and/or BKCI-SSH.

Table 1. Criteria for Inclusion in Database Queries

Note. Adapted from "D2.1 Systematic Review and Methodological Framework," in *H2020 Grant Agreement No 732332*, (p. 29), by D. Persico, C. Bailey, F. Dagnino, M. Haggis, F. Manganello, M. Passarell, and C. Perrotta, 2017, European Commission, located in City of Brussels, Belgium. Copyright © 2017, European Commission. Adaptation based on the review proposed in the work.

Keyword Category	Search Terms
Role-Playing Game-Related Keywords	RPG*, Tabletop roleplaying gam*, roleplaying videogam*, ttrpg*, mmorpg*, rpg*, crpg*
Game Mechanics Keywords	mechanic*, gameplay*, balanc*, physic*, control*, player movement*, rule*, design*, strateg*, interact*
Narrative Generation Keywords	storytell*, plot*, procedurally generated*, script*, character develop*, branching narrativ*, quest design*, narrativ* structure*, generative story*, dialogu* system*
Dynamic Interaction Keywords	real-time*, multiplayer*, social interact*, cooperative play*, networked play*, feedback loop*, AI interact*, adaptive gam*, non-player charact*, emergent behavior*

Table 2. Keywords Used in the Database Queries

Note. Adapted from "D2.1 Systematic Review and Methodological Framework," in *H2020 Grant Agreement No 732332*, (pp. 29–30), by D. Persico, C. Bailey, F. Dagnino, M. Haggis, F. Manganello, M. Passarell, and C. Perrotta, 2017, European Commission, located in City of Brussels, Belgium. Copyright © 2017, European Commission. Adaptation based on the review proposed in the work.

Database	Search String	Hyperlink
Web of Science	(TS=((“role-playing game*” OR “RPG*” OR “tabletop roleplaying gam*” OR “ttrpg*” OR “mmorpg*” OR “crpg*”) AND (“game mechanic*” OR “narrative generation” OR “dynamic interaction” OR “gameplay*”)) AND PY=(2018-2024) AND LA=(English OR Spanish))	WoS Link
SCOPUS	“role-playing game*” OR “RPG*” OR “tabletop roleplaying gam*” OR “ttrpg*” OR “mmorpg*” OR “crpg*” AND “game mechanic*” OR “narrative generation” OR “dynamic interaction” OR “gameplay*” AND PUBYEAR > 2017 AND PUBYEAR < 2025 AND (LIMIT-TO (LANGUAGE , “English”) OR LIMIT-TO (LANGUAGE , “Spanish”)) AND (LIMIT-TO (DOCTYPE , “ar”) OR LIMIT-TO (DOCTYPE , “ch”) OR LIMIT-TO (DOCTYPE, “re”) OR LIMIT-TO (DOCTYPE , “bk”)) AND (LIMIT-TO (SUBJAREA , “SOCI”) OR LIMIT-TO (SUBJAREA , “COMP”) OR LIMIT-TO (SUBJAREA , “ARTS”) OR LIMIT-TO (SUBJAREA , “PSYC”)) AND (LIMIT-TO (EXACTKEYWORD , “Human Computer Interaction”))	SCOPUS Link

Table 3. Search Strings Used in the Database Queries

Note. The results on SCOPUS were additionally limited to the subject areas of "Social Sciences, Computer Science, Arts and Humanities, Psycology," and the keyword was limited to "Human-Computer Interaction."

RESOURCES

Aarseth, E. J. (1997). *Cybertext: Perspectives on Ergodic Literature*.

Allen, F. (2016). *The Black Box Society: The Secret Algorithms That Control Money and Information* (Reprint edition). Harvard University Press. https://www.amazon.com/Black-Box-Society-Algorithms-Information/dp/0674970845

Antonius Alijoyo, F., Sneha Sri, S. S., Alapati, P. R., Yuldashev, D., & M, P. V. (2024). Ethical Considerations in Explainable AI: Balancing Transparency and User Privacy in English Language-based Virtual Assistants. *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 399–406. https://doi.org/10.1109/ICICV62344.2024.00069

Asimov, I. (1990). *The Foundation Novels 4-Book Set: Foundation/Foundation and Empire/Second Foundation/Foundation's Edge*. Bantam Spectra.

Bautista, Y. J. P., Theran, C., & Aló, R. (2024). Ethical Considerations of Generative AI: A Survey Exploring the Role of Decision Makers in the Loop. *Proceedings of the AAAI Symposium Series*, *3*(1), Article 1. https://doi.org/10.1609/aaaiss.v3i1.31243

Cheong, B. C. (2024). Transparency and Accountability in AI Systems: Safeguarding Wellbeing in the Age of Algorithmic Decision-Making. *Frontiers in Human Dynamics*, *6*. https://doi.org/10.3389/fhumd.2024.1421273

Dibbell, J. (1993). A Rape in Cyberspace, or How an Evil Clown, a Haitian Trickster Spirit, Two Wizards, and a Cast of Dozens Turned a Database into a Society. *The Village Voice*. http://www.juliandibbell.com/texts/bungle_vv.html

Faden, R. R., & Beauchamp, T. L. (with King, N. M. P.). (1986). *A History and Theory of Informed Consent* (1st ed.). Oxford University Press. https://global.oup.com/academic/product/a-history-and-theory-of-informed-consent- 9780195036862?cc=us&lang=en&

Fischer, G. (2012). Meta-Design and Cultures of Participation: Transformative Frameworks for the Design of Communication. *Proceedings of the 30th ACM International Conference on Design of Communication*, 137–138. https://doi.org/10.1145/2379057.2379083

Floridi, L. (2013). *The Ethics of Information*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199641321.003.0002

Fujibayashi, H., Hirano, H., & Tominaga, K. (Directors). (2017, March 3). *Zeruda no densetsu: Buresu obu za wairudo* [Action, Adventure, Fantasy]. Nintendo Entertainment Planning & Development (EPD).

Juul, J. (2005). *Half-Real: Video Games Between Real Rules and Fictional Worlds* (1st edition). Mit Pr.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.

Li, Y., & Zhu, J. (2024). An Ethical Study of Generative AI From the Actor-Network Theory Perspective. *International Journal on Cybernetics & Informatics*, *13*(1), 67–78. https://doi.org/10.5121/ijci.2024.130106

Munslow, J. (2024, April 29). I Tricked ChatGPT Into Being My Boyfriend. He Got Spicy Real Fast. *The Wall Street Journal (WSJ)*. https://www.wsj.com/lifestyle/chatgpt-ai-boyfriend- spicy-8ac6a6e9

Murray, J. H. (1998). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. MIT Press.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press. https://doi.org/10.2307/j.ctt1pwt9w5

Ryan, M.-L. (2003). *Narrative as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media (First Edition)* (1st ed.). Johns Hopkins University Press.

Ryan, M.-L. (2015). *Narrative as Virtual Reality 2: Revisiting Immersion and Interactivity in Literature and Electronic Media* (Illustrated edition). Johns Hopkins University Press.

Shaw, K., & Bryant-Monk, L. (2021, August 12). *TTRPG Safety Toolkit: A Quick Reference Guide V. 2.5*. https://drive.google.com/file/d/1M3LpDnVOc2G5UV03mWsqSU2QkDvHcmWX/view? usp=drive_link

Stavropoulos, J. (2013, July 31). *X-Card: Safety Tools for Simulations and Role-Playing Games*. http://tinyurl.com/x-card-rpg

Zhuk, A. (2024). Ethical implications of AI in the Metaverse. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00450-5