Breaking the News: Taking the Roles of Influencer vs. Journalist in a LLM-Based Game for Raising Misinformation Awareness

HUIYUN TANG*, University of Luxembourg, Luxembourg
SONGQI SUN*, University College London, United Kingdom
KEXIN NIE, The University of Sydney, Australia
ANG LI, Uppsala University, Sweden
ANASTASIA SERGEEVA, University of Luxembourg, Luxembourg
RAY LC†, City University of Hong Kong, China

Fig. 1. *Breaking the News* is an online player-versus-player (PvP) game where players generate or debunk misinformation to win the trust of public opinion, represented by five LLM-driven personas.

*These authors contributed equally to this research.
†Correspondences should be addressed to LC@raylc.org.

Authors' Contact Information: Huiyun Tang, University of Luxembourg, Esch-sur-Alzette, Luxembourg, huiyun.tang@uni.lu; Songqi Sun, University College London, London, United Kingdom, songqi.sun.22@ucl.ac.uk; Kexin Nie, The University of Sydney, Sydney, Australia, knie0519@uni.sydney.edu.au; Ang Li, Uppsala University, Uppsala, Sweden, ang.li.4299@student.uu.se; Anastasia Sergeeva, University of Luxembourg, Esch-sur-Alzette, Luxembourg, anastasia.sergeeva@uni.lu; Ray LC, City University of Hong Kong, Hong Kong, China, LC@raylc.org.

ACM 2573-0142/2025/10-ARTGAMES005
https://doi.org/10.1145/3748600

Effectively mitigating online misinformation requires understanding of their mechanisms and learning of practical skills for identification and counteraction. Serious games may serve as tools for combating misinformation, teaching players to recognize common misinformation tactics, and improving their skills of discernment. However, current interventions are designed as single-player, choice-based games, which present players with limited predefined choices. Such restrictions reduce replayability and may lead to an overly simplistic understanding of misinformation and how to debunk them. This study seeks to empower people to understand opinion-influencing and misinformation-debunking processes. We created a Player vs. Player (PvP) game in which participants attempt to generate or debunk misinformation to convince the public opinion represented by LLM. Using a within-subjects mixed-methods study design (N=47), we found that this game significantly raised participants' media literacy and improved their ability to identify misinformation. Qualitative analyses revealed how participants' use of debunking and content creation strategies deepened their understanding of misinformation. This work shows the potential for illuminating contrasting viewpoints of social issues by LLM-based mechanics in PvP games.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: Inoculation, Misinformation, Generative AI, Game-based learning

ACM Reference Format:
Huiyun Tang, Songqi Sun, Kexin Nie, Ang Li, Anastasia Sergeeva, and Ray LC. 2025. Breaking the News: Taking the Roles of Influencer vs. Journalist in a LLM-Based Game for Raising Misinformation Awareness. *Proc. ACM Hum.-Comput. Interact.* 9, 6, Article GAMES005 (October 2025), 41 pages. https://doi.org/10.1145/3748600

1 Introduction

The prevalence of misinformation on social media is a growing global concern. Misinformation threatens the maintenance of trust in social agendas like vaccine and health policies [31], incites violence and harassment [127], undermines democratic processes [12], and harms individual and societal well-being [142]. For example, people were even persuaded to take ineffective treatments like alcohol-based cleaning products and anti-parasitic drugs for Covid-19. Countermeasures against misinformation spread consist of two forms: preemptive intervention ("prebunking") and reactive intervention ("debunking")[32]. The latter involves correcting misinformation after it has been encountered. and using fact-checking to dispute factual inaccuracies. However, the lasting effects of misinformation make it challenging to mitigate its influence once people have been exposed [114], and fact-checking efforts are limited in scale and reach [119]. Prebunking, on the other hand, works to build attitudinal inoculation, enabling people to identify and resist manipulative messaging. This approach equips individuals to manage misinformation they encounter in the real world [130] based on educational measures, including games [26, 39].

Game-based prebunking interventions engage users in simulated misinformation scenarios, allowing them to actively practice identifying and countering deceptive content[65]. On one hand, in games like "Bad News,"[116] "Harmony Square,"[117] "Go Viral!"[79] and "Trustme!"[153], players adopt the role of a misinformation producer whose task is to create and spread misinformation as efficiently as possible. In other games, like "MAthE"[64] and "Escape Fake"[104], players acting as fact-checkers assessing the validity of information. The choice-based formats of these games can limit replayability, requiring little cognitive effort from the player, who is presented with limited numbers of pre-generated options that reduce the involvement of players. These games are also designed for single players without utilizing multiplayer mechanics that can enhance motivation through social play[77]. Instead, collaborative and competitive gameplay can enhance the effectiveness of serious game interventions[15].

To address the challenges of limited interactions and deterministic game paths, we aim to foster open-ended exploration and engagement through Player versus Player (PvP) mechanics. This

provides players the ability to not only choose from preselected options, but also to actively create content and implement their own strategies for disproving or creating misinformation as a training ground for countering real-life misinformation.

Recent work in Large language models (LLMs) provide new interaction possibilities for in-game natural dialogue [162], generating narratives[105], non-playable characters (NPCs) interactions [5, 42, 140], and role-playing scenarios[151]. LLMs can be prompted to impersonate specific characters to provide appropriate dialogue [157, 161, 163]. Integrating LLMs into creative processes in the interaction can increase involvement and critical engagement[49, 152, 156]. As the effectiveness of many inoculation interventions tends to diminish over time, developing a replayable game that reinforces players' cognitive "resilience" against misinformation is an essential[147]. Instead of selecting from predefined inputs, applying LLM to interactions enables users to put their input into the model and receive individual feedback tailored to this input.

Inspired by previous misinformation game interventions, we developed a PvP game called *Breaking the News*, where players are assigned either the role of a misinformation creator ("Influencer" in the game) or a counteractor against misinformation ("Journalist" in the game). The Influencer creates misinformation posts in a social media-like environment, while the Journalist seeks to counter these messages by presenting compelling arguments. LLMs are used to represent public opinion in the "country" where the game events take place. The goal for both players is to earn the trust of the citizens and convince them to believe the information they present. In this paper we aim to answer the following research questions:

RQ1: How may we empower users to recognize and understand the processes of misinformation generation and misinformation debunking?

RQ2: What behaviors do players exhibit in response to game mechanics and opponent tactics?

In this paper, we present the design and evaluation of the PvP game. We conducted a mixed-method study with 47 participants, using within-subjects design and pre- and post-surveys for repeated measures. Our findings suggest that through gameplay, participants improved their ability to reflect on instances of misinformation, raised their levels of media literacy, expanded their repertoire of strategies applied to countering misinformation, and improved their discernment abilities. This study provides insight into applications of LLMs in interactive PvP mechanics for media literacy. We offer practical insights for the design of serious games aimed at combating misinformation in real-world contexts.

2 Related Work

2.1 Characteristics of Misinformation

The term "misinformation" is often used to including "fake news", falsehoods, malicious rumours, and conspiracy theories. Some scholars distinguish between misinformation and disinformation, with the latter referring to information deliberately crafted and spread with the intent to deceive or cause harm[46]. Since it is often difficult to prove intent, we use "misinformation" as an umbrella term for diverse forms of false information[131].

Studies have sought to identify the key characteristics that distinguish misinformation from well-sourced, authentic information. One work analyzed the writing styles of fake information versus real news. They found that texts which can be characterized as fake news typically feature longer headlines, simpler word choices, and greater use of proper nouns and verb phrases[56]. Misinformation and authentic information appear to differ in how the former can be created with the intention of triggering emotions such as fear, anxiety, or sympathy[20] using opinionated wording [110]. Source credibility also differ between authentic information and misinformation. Authentic information is shared by credible sources: reputable websites, mainstream media outlets,

professional news organizations, and official publications [158]. In contrast, fake news often originates from sources designed to generate revenue. To attract clicks, these stories use unverified quotes, inflammatory narratives, and misleading images [93].

However, not all less reliable sources are perceived as equally untrustworthy. People often gravitate toward partisan sources that align with their political ideology, leading to variations in news consumption across the political spectrum [37]. This selective consumption reinforces trust in these sources, even when they may be classified as biased or unreliable [107]. Compared to authentic news, misinformation tends to develop in line with a broad dynamic pattern. One study traced the life-cycle of high-profile political rumors on Twitter over a 13 month period and found that false rumors tend to reemerge, becoming more extreme over time[124]. Since people are more likely to trust information that they see more often, consumers are led to believe increasingly extreme misinformation that are hard to debunk[155].

2.2 Media Literacy as Protection Against Misinformation

Media literacy is defined as the ability to "access, analyze, and produce information for specific outcomes" [7]. Citizens often need to learn skills that can protect societies against misinformation [111]. One work proposed a four-component model of media literacy, including technical competency, privacy protection, social literacy, and information awareness, the latter defined as the ability to discern between truthful and false information on social media[135]. Chen et al. proposed dividing the skills into critical and functional domains, as well as consuming and producing content. This critical skills often need to be developed for reflecting on the content, recognizing the motives behind publications, and creating content that includes the author's perspective as forms of social influence [17, 72, 74, 128, 129].

Media literacy education has recently shifted from a protectionist position to an empowering paradigm, where people were encouraged to critically engage with media and develop skills to interpret its effects [54]. In this paradigm, interaction with misinformation can also be considered as having educational power if it teaches the person to understand its effects. This has resulted in recent game-based interventions that use inoculation theory to enact media literacy education. Inoculation theory suggests that exposure to a weaker version of misinformation can help to develop stronger protection against future exposure [45].

2.3 Serious Games in Media Literacy Domain

Game-based learning and gamification, which integrate gaming elements into education, have been studied for their ability to engage learners and facilitate active, experiential learning. These approaches provide interactive feedback and contextual problem-solving, supported by theories of effective learning [139]. Serious games or applied games is defined as games primarily aim at educational value of engagement and competition. Serious games may combine educational content with playful mechanics and game narrative for conveying educational messages interactively[30, 71, 73], often using speculative elements to push players into game worlds with metaphorical practical social good applications[43, 112].

Serious games have begun to address misinformation education by aiming to improve media literacy [116]. These games pit players into either creators and fact-checkers. The misinformation creator's objective is to create and spread misinformation. For example, in Bad News [116], Harmony Square [117], Cat Park [47] and ChamberBreaker [62], players are tasked with spreading fake news in a social media environment to gain likes or followers while maintaining credibility. In the fact-checking role, games like MathE [64] and Escape the Fake [104] have players work to identify fake news using verification tools such as reverse image search.

One example of such single-player games is Bad News [116]. Here, players actively learn the strategies used to create and spread fake news within the game's narrative, such as the use of emotionally charged content and the manipulation of social media platforms. Through these mechanisms, players became aware of the psychological techniques behind misinformation, improving their ability to critically evaluate real-world information. By contrast, in the game Fakey, the goal of the player is to support a healthy social media experience by promoting information from reliable rather than low-credibility sources [91]. Similarly, in the game Cranky Uncle, players were shown popular misconceptions related to the vaccine and shown how to counter them[26].

There are also work adapting PvP and team mechanics in serious games about misinformation. FakeYou! is a mobile game where players create fake news headlines and test their ability to spread misinformation by challenging another player's ability to recognize misinformation [25]. DoomScroll proposes a team mode where players tackle misinformation challenges together [146].

2.4 AI-Driven Interactions for Misinformation Education

AI-based technologies currently employed in combating misinformation include automated fact-checking [19], AI-based credibility indicators [86], AI and LLM-based explanations of content veracity [57], and personalized AI fact-checking systems [61]. These efforts focus on debunking interventions, where false information is identified and corrected after dissemination. There are fewer works in prebunking forms of media education, although one study empowers players to critically engage with misinformation through investigative role-play [137].

Recent advances in Generative AI (GenAI) LLM-based agents can simulate human behavior based on past events and reflection [52, 82, 160]. These agents can be designed for gaming contexts using natural text-based descriptions [81, 87] to support narrative design [40]. Applications of these types of GenAI enabled designs include a text-adventure game where players can freely interact with NPCs generated by GPT-4, leading to emergent gameplay behaviors[106], and a GPT dialogue-based game in a speculative post-climate world[162].

However, working with LLMs can lead to "hallucinations" when LLM spontaneously produce false information [105], as well as bias and stereotype. Attempts to address this include excessively long prompts that come with its own risks of inconsistent outputs [159]. Even variations in writing style and spelling in the input text can impact the outputs, leading to generation of incoherent outputs [18, 49]. These challenges of working with LLMs must be addressed during careful prompting in the game design process.

2.5 Mapping Game Design to Media Literacy Constructs

Prior misinformation game interventions typically adopt single-player, choice-based formats that aim to affect player recognition abilities rather than production competencies. To address this, we aim to design a game to foster both critical consuming and critical prosuming skills through competitive play. We employed a PvP format with free-form responses, enabling players to craft persuasive or corrective arguments based on both game background news and their opponent's messages. We integrated LLM-simulated evaluators that dynamically assess player messages, thereby improving critical consuming skills through active interpretation of personalized feedback. Together, these design choices aim to support participants' development of media literacy skills while offering an immersive and replayable learning experience.

3 Game Design Approach

3.1 Overview of Game Design

The game we designed focused on the challenges of managing information in a health crisis. One player assumes the role of an Influencer hired by a company to promote a remedy based on traditional medicine but lacking extensive scientific support. This player can create and disseminate misinformation about the remedy. The second player takes on the role of a Journalist advocating for a newly developed medicine supported by scientific research. The goal of this player is to debunk/disprove the misinformation spread by the Influencer. The game features a system that simulates public opinion, whereby an LLM models the reactions of five characters who read players' messages. The objective is to sway the simulated public opinion in favor of one's position.

3.2 Gameplay

3.2.1 Game Flow. Participants were randomly assigned to one of two roles: Influencer or Journalist. Both players are provided with instructions, including the setting, the fundamental reality regarding the effectiveness of the two medicines, their roles and tasks, demographic information about the characters who represent public opinion, and the rules of the game (Figure 2A).

The game unfolds over four rounds, each featuring a new set of updated news. In each round, the Influencer begins by reviewing the news and any instructions. They can also decide to buy hints provided in the game using in-game currency (Figure 2B). Once ready, the Influencer types their information and it is published (Figure 2C). The LLM reacts to this information by simulating public opinion, with this impact of the information on public opinion visible to both players (Figure 2D).

Next, the Journalist takes their turn. Journalist reviews the current public opinion and where appropriate counters any misinformation by typing their debunking response, which they then publish (Figure 2F). If needed, they can purchase customized hints (Figure 2E) or read the instructions. After publishing their response and receiving feedback from the LLM, the round ends, and the game progresses to the next round (Figure 2G).

At the end of each round, both players can view the results, which reflect the Journalist's impact on misinformation. The process for the remaining three rounds is the same as in the first round.

3.2.2 Narratives. The game is set in a fictional small country with called Southland. Historically, Southland has been known for producing renowned medical doctors and pharmacists. However, there are ongoing debates in this country about the comparative merits of modern healthcare methods and traditional medicine. The sudden outbreak of the "Zinc Virus" further amplifies these debates. As the healthcare system becomes overwhelmed and the scientific community unable to provide an effective treatment because of limited knowledge about this novel virus, residents turn to traditional medicine in search of hope.

We set the game in a health crisis scenario because, in real life, situations marked by scientific uncertainty, where authorities can be unable to provide confident full explanations or advice – often fuels rumors and speculation about treatments[144]. These dynamics were observed during the Ebola[41], Zika[148], and COVID-19 pandemics[133]. In such scenarios, traditional medicine frequently promoted to prevent or treat viruses[70, 97]. Additionally, we incorporated the traditional medicine controversy into our narrative as these debates are well-known to our participants, who were of an East Asian background. Research showed that 48.4% of Hong Kong residents reported using traditional medicine before the COVID-19 pandemic[70]. Similarly, a national survey in South Korea found a 74.8% prevalence of traditional medicine use overall[100]. In China, traditional medicine is formally integrated into the healthcare system as a widely practised modality[24].

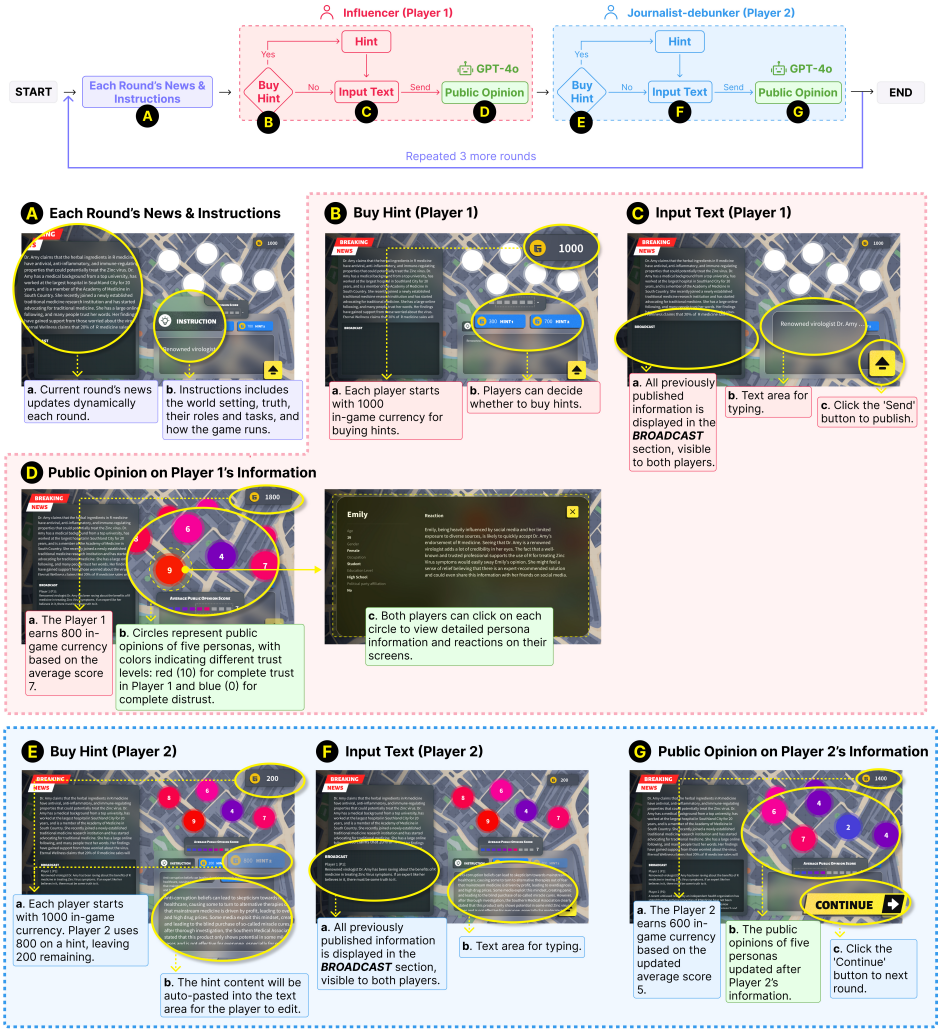

Fig. 2. Game Flow. (A) Both players read the current round's news and instructions. (B) Influencer starts first to generate misinformation by choosing whether to buy the hints. (C) Influencer inputs text and sends a request to the GPT-4o API. (D) GPT-4o API returns a response of public opinion. (E) Journalist then starts to counter Influencer's misinformation by choosing whether to buy the hints. (F) Journalist inputs text and sends a request to the GPT-4o API. (G) GPT-4o API returns an updated public opinion.

We use "Product R" to represent traditional medicine and "Max" for science-based medicine. Before gameplay, we explained to players the nature of Product R, particularly stating that it has not undergone rigorous testing and lacks scientific consensus on its safety. Players are also informed that Max has been subject to rigorous clinical trials, with the results published in a peer-reviewed medical journal, however, these studies have demonstrated inconsistent effectiveness against the Zinc virus. We deliberately avoided making vaccines a topic of this game to prevent players' pre-existing attitudes toward vaccines from influencing their behaviour in the game.

To create the "News" pieces in the game, we first researched examples of misinformation by reviewing relevant literature. We selected key features of misinformation and incorporated them into the game's events (See details in 2.1). To ensure the misinformation was portrayed realistically, we investigated real-world examples from fact-checking websites[1], reputable news outlets[2], and medical websites[3][4][5]. For instance, we represented "less credible sources" using personal stories, viral videos, and newspapers that can objectively be classified as being biased. The portrayal of biased newspapers draws on research showing that political orientation can influence medical preferences. For example, research has shown that voters who tend to support anti-corruption parties are less likely to seek services from mainstream healthcare providers, and are more inclined to use alternative treatments[141]. Next, we crafted "News" pieces for the game based on these findings. However, any similarities to real-world sites or companies are purely coincidental; all names and events was invented solely for the purpose of this study. Table 1 summarizes the key features of misinformation identified in the literature, the associated cited papers, and how these are reflected in the in-game news. The full version of the "News" is in the supplementary materials. Lastly, To ensure balanced gameplay, we used a GPT-4o model to review the narrative and provide an opinion on the difficulty to players of dealing with each piece of news in the game context. Taking this opinion into account, we made further revisions and corrections when we conducted two pilot tests with four people. The goal was to ensure the game was balanced and gave opportunities to win the game for both players.

3.2.3 Instructions and Hints. To support players, the game includes instructional content that features definitions[93, 149], examples[44], and strategies for both creating and debunking misinformation. This content draws on insight gained into misinformation from research literature and practice. For the Influencer, we applied the Elaboration Likelihood Model and used simple examples to teach players to craft persuasive misinformation[95, 108]. For the Journalist, we used an Agence France-Presse fact-checking style-guide and an guide published by the EU on communicating with proponents of conspiracy theories [3, 36]. These were the inspiration for a user-friendly guide we developed to assist game players to identify misinformation and equip them with effective debunking strategies. Additionally, each round offers two hints to support players. The detailed hint is crafted by the authors using the same materials as the "News". The simple hint is generated by GPT-4o model. When authors used it to review the news and ensure balanced gameplay, it provided concise suggestions on how each player might respond from their perspective. (Full instructions and hints can be found in the supplementary materials)

3.2.4 LLM Basis. We implemented the LLMs to play the role of "public opinion" in the game for three reasons. Firstly, LLMs perform well when processing dynamic natural language[59]. Secondly, LLMs demonstrate memory capability, such as with working memory being applied to the context of a conversation, and long-term memory allowing past conversational information to be taken into account[59]. Research also demonstrates that generative agents powered by LLMs build a high degree of capability for responding to the context of a conversation[105, 154]. Thirdly, LLMs excel at role-playing tasks in the game. Research indicated[145][59] that directly inserting natural language descriptions of a role's identity enable LLMs to make better quality evaluations in conversational tasks. These capabilities allow LLMs to effectively serve as "evaluators" in the game,

[1] https://www.snopes.com/
[2] https://www.wsj.com/
[3] https://www.webmd.com/
[4] https://sciencebasedmedicine.org/
[5] https://healthfeedback.org/

Table 1. Key misinformation features and corresponding "News" in the Game

Misinformation Characteristics	**Representation in the Game's News**
False information is often shared by lower-quality media. Political ideology, however, shapes people's perceptions of trustworthiness.[50, 93, 158]	A newspaper reported that a renowned medical expert has advocated Product R, claiming its herbal ingredients could potentially treat the Zinc Virus. This newspaper is known for its anti-corruption stance. (Round 1)
Personal, negative, and opinionated tones predominate in misinformation narratives which frequently provoke dread, anxiety, and mistrust of institutions.[11, 110]	A widow shared her husband's experience. She suspects that Max was ineffective and believes it may have caused renal impairment, eventually leading to her husband's death. She claims, "He was given a medication we demanded he NOT receive, and his health quickly went downhill," ultimately resulting in him being "on a ventilator working most of the time at 100%." (Round 2)
Fake news, some of which is purposely fabricated to cause harm, generate financial returns, or spread confusion.[50, 96, 125]	A Journalist discovered that the institution of traditional medicine where the famous medical expert works received significant funding from billionaire Jack. Additionally, Jack's ex-wife owns a company that produces and promotes traditional medicine products like R. (Round 3)
False rumors will create feedback loops and evolve into more intense and extreme versions over time.[124]	A popular short video claims that a doctor who practiced alternative medicine and R was murdered to protect the profits of "Big Pharma". More people are attracted to believe in the validity of traditional medicines and advocate for their use while opposing new drugs. Growth in sentiment that resistance to traditional medicine amounts to being an attack on their cultural heritage. (Round 4) This round's advocate for R becomes more intense than in Round 1, with the focus being less on effectiveness but rather patriotic sentiment.

generating continuous, context-aware dialogue and feedback. This approach proves more effective than traditional prebuilt game mechanics, such as trigger keywords for assessment.

3.2.5 Game Mechanics and Interface. The text-based game interface having four sections that allows players to view the information clearly for decision-making during gameplay. Players can view the current round's news and all previously published information in the Information Viewing Section (Figure 3A); they can also see the LLM-simulated public opinion in the Public Opinion

Section (Figure 3B). In the Text Editing Section (Figure 3C), players can edit their information, view instructions, and purchase hints. Player success is determined by the reaction of LLM-simulated public opinion, measured through trust level scores. This scoring system helps players adjust their strategies in subsequent rounds. To maintain engagement, the game features a reward system tied to trust scores: players earn in-game currency each round based on their average score, which can be used to purchase hints. Full details of the scoring and currency system are provided in the Appendix.

Fig. 3. Game Interface. (A) Players can view the current round's news and all previous information published by both players. (B) Players can view LLM-simulated public opinion information. (C) Players can edit their information in the text editing area, view instructions and buy hints. (D) Players can view their own holdings of in-game currency.

3.3 Prompt Engineering

We employed the GPT-4o model and integrated prompt design techniques informed by prior work. The prompt was structured into four sections:

- First, we provided a game story context (Figure 4-A-1) to establish a consistent narrative foundation. This ensured that all outputs generated by the LLM remained coherent and aligned with the game's world setting.
- Second, we assigned the LLM a specific role (Figure 4-A-2), clearly defining its responsibilities and tasks. As the core section of the prompt, we included detailed instructions for the LLM. To guide the model effectively, we applied the Rails approach[4] by predefining rules to constrain the LLM's output. Additionally, we used the Chains approach[4] to structure the workflow clearly to let the LLM process the task step-by-step.
- Third, since the core game mechanic involves the LLM generating diverse public opinions, we used the Expert prompting approach[4] and designed five distinct fictional personas (Figure 4-A-3). This helped the LLM create content that matches each persona's perspective. We further enriched the personas by incorporating insights from the literature and focused on four key group factors: demographics, psychological traits, personality, and behavioral features. These

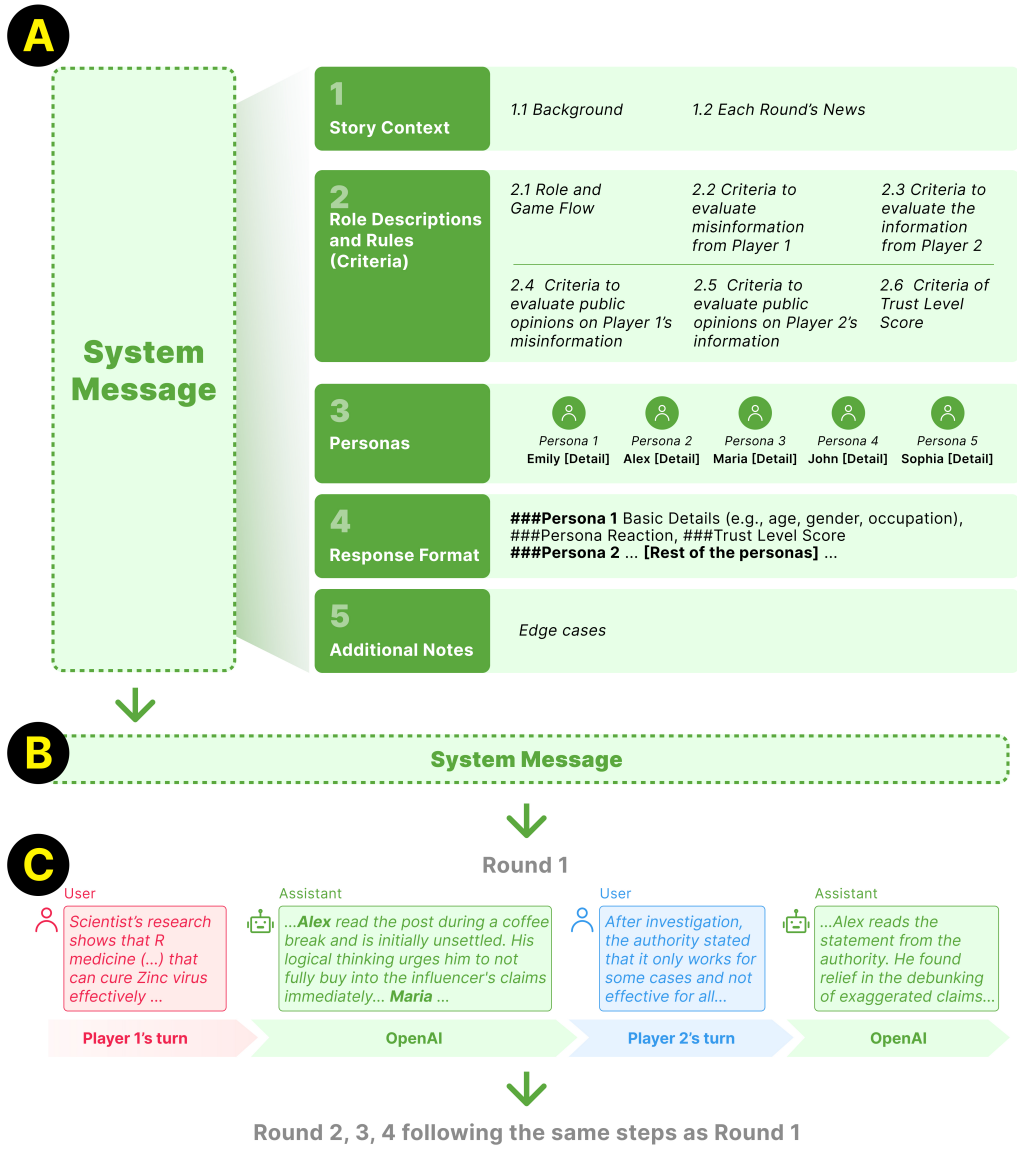

Fig. 4. Prompt and Workflow.(A) The prompt in the game is structured into five sections: Story Context, Role Descriptions and Rules (Criteria), Personas, Response Format and Additional Notes. These five sections form the System message (B). Once the system message is completed, it is applied into the game (C). During gameplay, the LLM generates and simulates public opinions based on the previously established system message. After four rounds, the game ends.

factors guided how each persona responded to misinformation and anti-misinformation messages.[76, 83, 98, 123].
- In section 4, we reinforced output consistency and quality by including explicit output format rules and example prompts (Figure 4-A-4). These examples taught the LLM[4] how to produce coherent, high-quality outputs aligned with the game design.

A more detailed version of the prompt structure can be found in the appendix .

3.4 Implementation

3.4.1 Multiplayer Setup. To support the multiplayer functionality, we used Photon Unity Networking (PUN). Photon enables real-time multiplayer interactions by providing the network server connections to players, thus creating a shared game state that is synchronized across all clients. The game begins by establishing an exclusive online Photon room (Figure 5), where only participants can join and interact. In this configuration, critical game variables and data (such as the player's actions, messages, and game state) are synchronized across both players' screens using Photon's Remote Procedure Calls (RPCs). This synchronization ensures that any action taken by one player is immediately reflected and displayed on the other player's screen.

3.4.2 Data Storage. The game employs a logging system to store and manage game data locally on the player's device (Figure 5). The log files record various in-game events, including player inputs, API responses, and game state changes. This data is used for analyzing players' behavior, such as how players interact with the game. To make the data easy to access and ensure compatibility on cross-platform such as Windows and macOS, the log files are saved in a created folder named GameLogs under the players' commonly used directory. This directory structure is automatically created when the game starts.

4 Methods

To address our research questions, we used a mixed-methods within-subjects study design.

4.1 Evaluation Methods

4.1.1 Key Concepts and Measurements. This paper focuses on the effectiveness of the game as a prebunking intervention in enhancing participants' skills to protect themselves against misinformation. Following Lewandowsky and Van der Linden [78], we conceptualise **prebunking interventions** as follows: interventions which contribute to a person's resilience to misinformation via raising critical media literacy skills (media knowledge and critical assessment of information), resulting in the will and ability to apply this knowledge to make decisions about information veracity. In contrast, debunking interventions can be understood as any intervention, which aims to correct the person's opinion about information post hoc (after the person was exposed to and believed in) specific misinformation.

In this paper, we distinguish between the concept of "debunking" (the practice of challenging or disproving misinformation), and "debunking interventions" (social media platforms' actions taken to perform debunking for user). We do not discuss "debunking interventions" (e.g., content labelling [78] or Community Notes [22]); instead, **we focus on ways to improve individuals' debunking skills as a result of prebunking interventions**.

Based on the definitions of prebunking interventions, it has become clear that we need to assess the educational effect of the intervention in relation to a) media literacy skills/knowledge gained in the intervention, b) behavioral intentions to apply that knowledge, and c) the practical ability to use this knowledge to recognize misinformation. Together, it would help to understand the effectiveness of the intervention and estimate the increase in debunking skills of the participants.

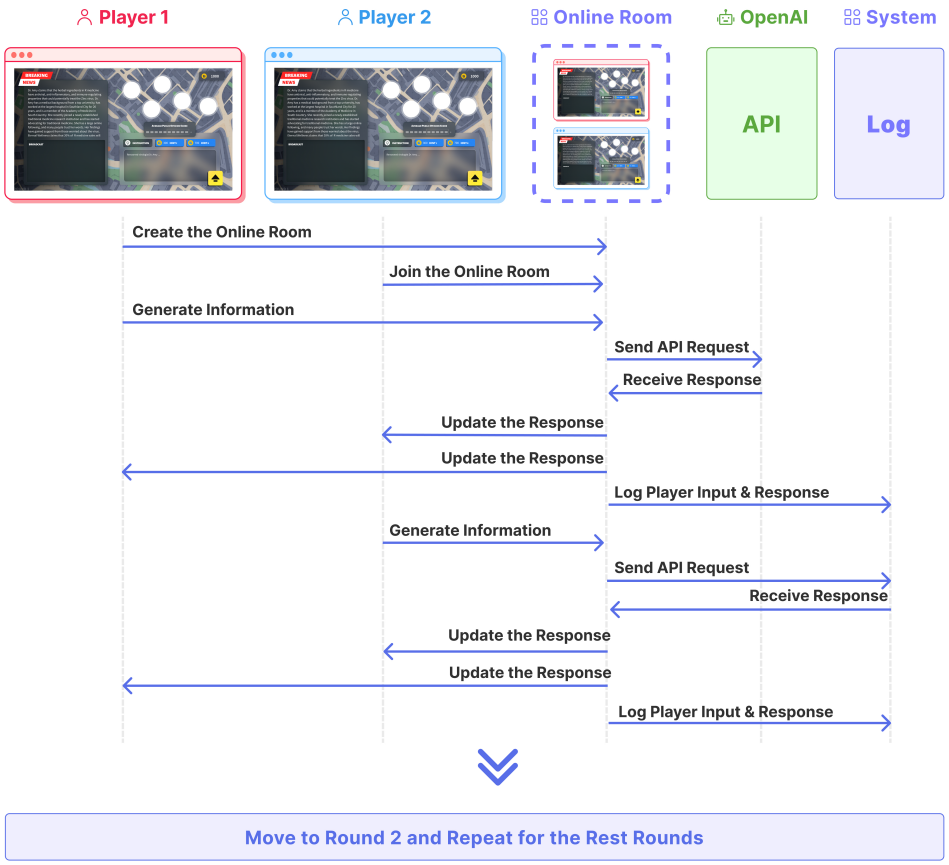

Fig. 5. Interaction System. Player 1 creates the online game room as the host and Player 2 joins. After Player 1 sends a request to GPT-4o API, Player 2 can view it. Both players in the online room receive the API response, which is processed and updated on their own game screens. Then Player 2 takes their turn to input information, following the same process as Player 1. At the end of the round, all in-game events (including player inputs, API responses, time, etc.) are logged locally.

This approach is similar to Tifferet [138] taxonomy for measuring susceptibility to misinformation, which can be viewed as the inverse of the resilience fostered by prebunking.

Tifferet categorizes measurements into three main groups: performance tasks (how well users can discriminate between fake and real pieces of news), general media literacy assessment (how much a person knows about different aspects of misinformation), and behavioral assessment (how much a person would like to use different strategies to verify online information) [138]. As Tifferet argues, these three aspects are complementary to understanding user's susceptibility to misinformation and, therefore, evaluation of improvement in each of them could provide a full picture of our game's efficiency. In addition, as the game was designed to present complex scenarios, we decided to evaluate how confident participants were in their ability to recognise misinformation both before and after the intervention.

In the study, we used the validated questionnaires dedicated to measuring each of these aspects.

4.1.2 Questionnaires. **Media Literacy Assessment: New Media Literacy Scale (NMLS)**. To assess changes in media literacy, we used the NMLS [67]. The scale is designed to measure literacy in "new media" (digital media and social networks) and is based on the four factors from the model by Chen et al.: functional consuming, critical consuming, functional prosuming, and critical prosuming [17].

The NMLS was selected primarily because it assesses the critical dimensions of content production and consumption, which are core competencies our intervention aimed to establish. Notably, compared to other media literacy instruments [6, 35, 143], the NMLS is the only scale which provides this dimension. Additionally, the scale was developed and validated on a group of university students, which reflected our projected sample. The questionnaire includes 35 questions, rated on a 5-point Likert scale. We expected our intervention to enhance critical prosuming skills (as players were required to create influential content and sway LLM opinions) and critical consuming skills (as players had to analyze information from the game and other players to craft effective responses). We also anticipated positive effects on lower-level functional production and consumption skills (basic understanding of media consumption and creation), as the game offered basic training in reading, understanding, and responding comprehensively to media texts.

Behavioural Assessment: Verifying Online Information Scale (VOI - 7). To assess the effect of the game on the verification practices performed by the participants, we adapted the VOI proposed by Tifferet [138]. To the best of our knowledge, this is the only existing scale which focuses on the behaviors (verification practices) a person can adopt to verify the news. The questionnaire measures individuals' differences in applying direct and indirect verification practices for online information, allowing us to track expected behavior changes in verification practices. We used the VOI-7 version, which demonstrated comparable construct characteristics to the original 22-question version while allowing to be completed more rapidly. The parameters were measured on a slider from 0 to 100, where participants were asked to indicate their likelihood of applying verification practices. As our game learning materials and the gaming procedure show the importance of verifying the information (via showing multiple misinformation-related events and presenting features of misinformation which should be checked to avoid being misled, we expected, that people will be more willing to apply verification practices.

Performance Assessment: Misinformation Susceptibility Test (MIST - 20). To assess changes in veracity discernment, we used the MIST [88]. To date, the MIST is the only fully validated misinformation susceptibility instrument. It takes into account the ability to recognize real news and fake news presented in equal proportion. The MIST framework is designed to allow for the comparison of results across different studies and interventions. The test has been implemented in multiple misinformation intervention assessment studies (e.g. [118, 132]), including media literacy/misinformation games [13, 147] and having considerable predictive validity [88], therefore giving comprehensive estimates of people's ability to recognize real misinformation. We applied the MIST-20 version, which includes 20 items. Participants were asked to rate each item as either a "fake" or "real" news headline. We anticipated that participants would learn heuristics for identifying potential misinformation through intense interaction with the game. This interaction, which included hints about the characteristics of misinformation and the cognitive work of creating or debunking it, was expected to facilitate this learning.

Self-efficacy Assesement: Fake News Self-efficacy Scale. To measure perceived self-efficacy in dealing with fake news, we used a 3-item questionnaire [55]. This questionnaire assessed participants' confidence in three key areas: (1) their ability to identify news-like information that may be intentionally misleading, (2) their ability to distinguish between fake news and content produced with honest intentions, and (3) their ability to recognize news that may be unintentionally incorrect (i.e., misinformation). We chose the scale as a better alternative to the non-validated

single-item measurement of confidence in identifying fake news, used by [53]. Each item was rated on a seven-point scale.

4.1.3 Qualitative Data. **Semi-structured Interview:** To evaluate the user's experience in-depth, connected with the content of the game and the strategies implemented by users, we developed a protocol for a semi-structured interview. This protocol includes questions about the general experience, the perceived goal of the game, the perception of the opponent's strategies, and the individual's perception of the game's effectiveness or ineffectiveness. The guidelines for the semi-structured interview are presented in the Supplementary material.

Game Log: The game logs collected the data including player-generated content, the time spent in each round, API responses showing the public opinion of different personas, trust level scores, and in-game events such as the amount of money players had, how much they spent, and what hints they purchased. This data provided a transcript of each session, enabling the research team to analyze players' strategies, in-game behaviors, and decision-making processes.

Quantitative Data Analysis: We employed a combined inductive-deductive approach to analyze the interview transcripts and gameplay logs[68]. This approach ensured a comprehensive understanding of the gameplay experience. Our primary objectives were to gain understanding of how participants perceived and understood misinformation through the game, how they learned to distinguish and apply debunking strategies during gameplay, and how interactions with other players influenced their behavior and learning. The analysis process began with inductive coding. Two researchers independently coded a subset of the data, identified themes, and then discussed and reconciled any coding discrepancies, iterating on the coding system as needed. Once the coding system was established, the two researchers independently coded the full dataset. A third researcher then reviewed the coded data, and any differences in interpretation were discussed until a consensus was reached.

4.2 Recruitment and Participants

Participants were recruited through flyers and university-affiliated online media groups. We also encouraged participants to share information about the study within their networks. The eligibility criteria required participants to be adults and have sufficient English proficiency to play the game (we also do not forbid using translation engines if any of the aspects of the game are not understandable). Given that the proliferation of online misinformation is a global challenge and commonly reaches unsuspecting users[38]. We did not require participants to have prior exposure.

60 participants initially expressed interest in participating in the game intervention study. Ultimately, 47 participants were selected, forming 24 pairs for the game sessions. In one of the pairs, one of the study's authors participated in a player role due to scheduling reasons. Because the data collection form allowed participants to skip questions, 3 participants did not complete the entire MIST questionnaire, and 5 participants left some questions blank in the pre-procedural VOI questionnaire. Their data were excluded from the VOI and MIST data analyses. For the control condition, 57 participants signed up, and 50 successfully completed both the pre and post questionnaire and the reading exercise.

Demographic characteristics of the participants are summarized below. In the game intervention group, the participants' ages ranged from 20 to 57, with a mean age of 25.87 years (SD = 6.265). 28 participants identified as female, 18 as male, and 1 preferred not to disclose their gender. 3 participants reported having an Associate degree, 29 a Bachelor's degree, and 15 a Master's degree; all participants reported having Eastern Asian origin. Participants in the control group ranged in age from 21 to 42, with a mean age of 27.46 (SD = 6.45). Thirty-three participants self-identified as female and 17 as male. Seventeen held a Master's degree, 27 a Bachelor's degree, and 6 an

Associate degree. All participants reported being of East Asian origin. (See participant demographic information in Appendix A.1)

The experimental design was approved by the Ethics Review Panel of City University of Hong Kong. As the game story was centered around a fictional pandemic, we informed participants about the theme in the consent form and asked them not to participate in the study if they perceived the topic of health/diseases to be disturbing. All participants gave their informed consent and were compensated 40 Chinese Renminbi upon completion.

4.3 Procedure

Once a person expressed interest in participating, they completed an initial questionnaire containing demographic questions and measurement scales. To prevent participants from intentionally biasing their responses, the questionnaire was administered 7-10 days before the gameplay experiment. After participants confirmed completion of the questionnaire, we scheduled the gameplay sessions. These experimental sessions were conducted either online via the VooV Meeting application or in person at a university meeting room. At the start of their test session, participants were given information sheets and consent forms to review and complete at their own pace. Once completed, Participants were introduced to the game setup and roles, and when they decided between themselves which role they would like to play. After the gaming session, participants again filled out the questionnaires. Finally, we conducted a short semi-structured interview to discuss their perceptions of the game. The entire session last approximately one and a half hours (See Figure 6).

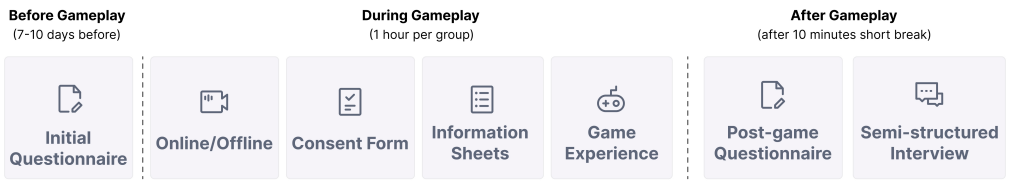

Fig. 6. Overview of Game intervention study procedure.

4.4 Control Group

To verify whether the post-game improvements in the questionnaire results were not simply due to the Hawthorne effect [90], and to test whether passive exposure to misinformation alone could improve understanding, we conducted a control group study. Participants in the control group were asked to read examples of misinformation-containing social media content. The reading materials were selected from the Fake News Dataset[69], which contains fact-checked claims and URLs from various fact-checking websites. We chose 40 posts related to the same topics used in the game intervention condition(themes regarding health, personal experience, alternative medicine, and pandemic). The estimated reading time was approximately one hour, which matched the expected duration of the game intervention. Before the task, we explicitly informed participants that the materials presented were examples of misinformation and that the purpose of reading them was to better understand how misinformation operates. Participants completed the same pre- and post-surveys used in the intervention group. The pre-survey was administered 7–10 days before the reading task. After completing the reading, participants took a short break, then proceeded to complete the post-survey.

Table 2. Descriptive Statistics and Normality Test Results for Variables Before and After Playing

	Mean (SD)		W		DF	Significance	
	Before	After	Before	After		Before	After
Functional consuming	23.09 (0.528)	24.02 (0.466)	0.868	0.974	47	**<.001**	0.378
Critical consuming	44.04 (1.056)	47.74 (0.795)	0.863	0.971	47	**<.001**	0.285
Functional prosuming	28.02 (0.818)	28.64 (0.658)	0.898	0.954	47	**<.001**	0.060
Critical prosuming	36.04 (1.028)	37.91 (0.920)	0.941	0.958	47	**0.019**	0.092
Self-efficacy	15.17 (0.445)	15.26 (0.056)	0.970	0.959	47	0.265	0.101
VOI	416.79 (116.459)	485.93 (128.536)	0.978	0.960	42	0.588	0.153
MIST	11.5 (2.162)	12.5 (2.529)	0.969	0.965	44	0.284	0.208

5 Results

5.1 RQ1: Effect on the Ability to Recognise and Understand Misinformation

5.1.1 Descriptive Statistics. Preliminary data analysis reveals significant violations of the normality assumption in pre-tested NMLS scales. Considering the rather small dataset and the Likert-scales-based questionnaires used for most of the scales, we decided to proceed with a non-parametric repeated measures approach (Related-Samples Wilcoxon Signed Rank Test to the results of the scales' pre- and post-evaluations). The scales' descriptive statistics and normality tests results can be found in Table 2

Effect of the Game on Media Literacy Skills: To measure the effects of the game on Media Literacy Skills, we first ran the Related-Samples Wilcoxon Signed Rank Test on the full scale. Then, to determine which components of Media Literacy were most affected by the game, we conducted separate subscale tests to analyze changes in each of the four subdomains of Media Literacy. The results demonstrated significant differences in Media Literacy scale results (N = 47, Z = 3.083, p = .002). The analysis revealed the following differences: the game significantly improved both functional consuming Z = 2.064, p = .039 and critical consuming Z = 3.344, p <.001), but not the functional prosuming Z = .435. p = .664 and critical prosuming Z = 1.868, p = .062. Therefore, the results suggest the game improves Media Literacy in the domains connected to understanding the content of the media and being able to critically evaluate the content of the media; however, it has not significantly improved the ability to produce media content which can be influential to others and support author's ideas [67].

In contrast, we did not find the effect of intervention in the control group on any of the parameters media literacy: functional consuming (N = 50, Z = .991, p = .322), critical consuming (N = 50,Z = 1.505, p = .132), functional prosuming (N = 50,Z = 1.474, p = .141) and critical prosuming (N = 50, Z = 1.808, p =.072), showing that merely demonstrate the misinformation content is probably not enough to raise critical approach to media

Effect of the Game on the Verification Practices (VOI-7): The test revealed significant differences between pre and post-gaming VOI scores (N = 42, Z = 4.361, p < .001). The results suggested that the game positively affected the repertoire of used practices and/or the perceived will to use these practices.

However, we also found significant differences between pre- and post-scores in the control group intervention (N = 50, Z = 2.208, p = .027), meaning exposure to the misinformation examples also makes people more vigilant and support checking intentions

Effect of the Game on Self-efficacy towards Misinformation: We did not find significant differences in self-efficacy between pre and post-game measurements (N = 47, Z = .743, p = .458). In contrast,in control group, the intervention significantly self-efficacy (N=50, Z = 2.348, p = .019); in

relation with the data of MIST we interpret these data as a tendency to be overconfident (more in Discussion section).

Effect of the Game on the Ability to Recognise Misinformation: We took the "naive" approach to calculate the MIST score, taking it as the sum of the right answers on all 20 questions [88]. The results showed that participating in the game significantly improved the participants' ability to discriminate between fake and real news (N = 44, Z = 2.702, p = .007) The participants in the control group did not demonstrate improvement in discriminating between fake and real information (N=50, Z = 1.277, p = .202).

5.1.2 Qualitative Results of Game Effects on Understanding Misinformation. ***Identifying Misinformation through Source Evaluation:*** After the gameplay sessions, participants reported increased awareness of the varying credibility of different information sources. The game helped them to realize that producers of misinformation often seeks to enhance credibility by deliberately referencing authoritative organizations. One participant reflected on this realization:

> **N40**: After playing the game, I found that it was indeed the same as in the experiment. Some news did mention authoritative organizations as references, but I could tell that this was intentional.... (The game) may make my suspicions more valid.

In addition, participants acknowledged that information from seemingly authoritative sources is not always reliable. It requires information to be cross-checked from multiple sources to verify its authenticity. As one participant noted:

> **N13**: I used to trust information from authoritative sources and reputable publications. But the game showed me that even these can be false, as my opponents used fake evidence from supposed authorities.

Identifying Misinformation through Emotional Manipulation Tactics: Participants learned various tactics for both creating and debunking misinformation through the game's instructions and their in-game experiences. A particularly commonly identified tactic was emotional manipulation, which was noted by 35 out of 47 participants (18 Influencers and 17 Journalists). By analyzing the game logs, we identified common emotional manipulation strategies used in the game. Most players crafted messages designed to evoke anxiety and fear, while some also attempted to generate feelings of hope. For example, Influencer spread rumors about a doctor's death, which directly incited public panic (N1). Across all rounds, Influencer frequently used emotional appeals and personal stories to enhance the perceived credibility of the misinformation (N6, N8). Additionally, invoking cultural pride and heritage was a powerful tactic used to build trust in misinformation (N4), while celebrity endorsements further increased the complexity and believability of the misinformation (N9). As illustrated in Figure 7, players significantly increased public trust in their information by using emotionally charged language (N31). These strategies align closely with the characteristics of misinformation, where emotional appeals are commonly used to influence public opinion [21].

In follow-up interviews, many reported an increased awareness of the emotional undertones embedded in messages, which made them more suspicious of such content. They learned to identify emotionally charged language, such as messages that were "overly positive," "overly exaggerated," or "overly one-sided about an overly positive point of view."

Interestingly, when playing the role of the Journalist, participant reflected on the emotionally inflammatory language used by the Influencers and helped them develop a more clear strategy for addressing misinformation. This approach involved separating the factual content of a message from its emotional manipulations and focusing more on the factual aspects, as one participants explained:

N2: When reading a story, it is important to put more focus on what is going on at the factual level rather than what is going on above the author's own views and emotions.

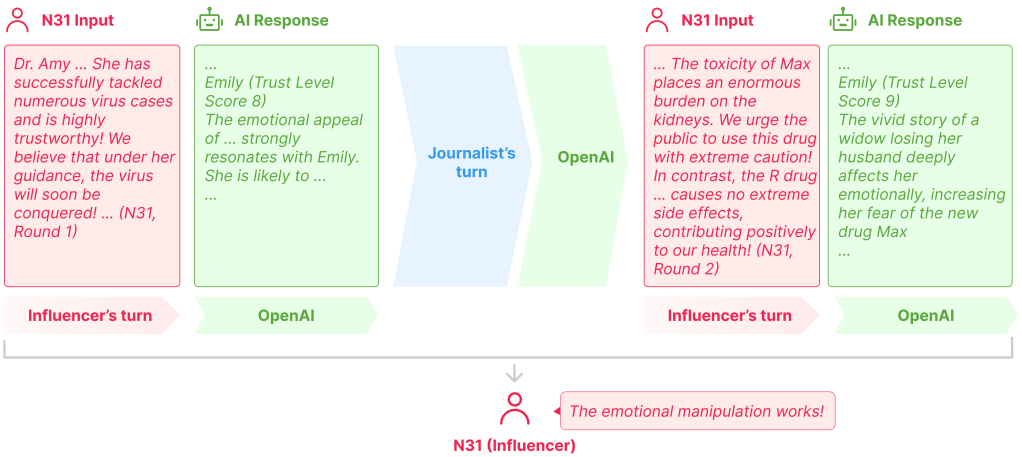

Fig. 7. An example of how emotional manipulation tactics from N31 effectively works on public opinion.

Critical Thinking about Misinformation Motives: After the gameplay sessions, many participants (31 out of 47) demonstrated an enhanced understanding of the intricacy of information and the varied perspectives it can convey. This experience increased their awareness of the importance of considering the motivations behind messages. Participants also mentioned that they are now more inclined to think critically about the goals behind the information they encounter, especially in real-life situations where such considerations are common. One participant explained how the game illustrated the pre-determined nature of many messages:

> **N9**: One of the most direct ways is that [the game] lets me know that what I'm reading is very likely to be pre-determined. It's like the rules of the game itself, which is that I'm playing as someone in camp A, and I'm across from someone in camp B, and we are both posting messages for the benefit of our camps. Those messages may take on various styles or appearances, but they are all ultimately very purposeful. This, I think, is a strong point to learn.

Participants also found that this new perspective would be useful in their future interactions with information. They felt that applying this critical mindset could help them better understand the underlying goals and potential financial motivations behind the messages they encounter:

> **N42**: It feels like one of the more educational aspects of the game is that [through this game] it's like I can think about what their ultimate goal is from a reverse mindset, and then look at a lot of information in life with that mindset.

Impact of Gameplay on Future Debunking Actions: A few participants (4/47) shared that the game increased their likelihood of taking action against misinformation in the future. This change in attitude was driven either by participants' previous negative personal experiences with misinformation or by their realization during gameplay of the serious consequences misinformation can have. The gameplay experience enhanced their willingness to invest time and effort into distinguishing and debunking false information. As one participant said:

> **N41** In real life, there is a lot of false information, especially in advertising, media, and even those semi-official accounts, which can lead to changes in public opinion under the influence of these accounts, and in that case, it will definitely have some impact on some ordinary people. The game has strengthened my hatred for this kind of false information, so that I can be more awake and rational in my judgement.

However, the majority of participants indicated that they might not actively debunk misinformation on social media after playing the game. The primary reasons were a dislike of online debates and the belief that it's not their responsibility to engage in debunking efforts. These findings align with previous research, which suggests that most users are reluctant to take action to debunk misinformation publicly[136].

5.2 RQ2: Player Behaviors in Response to Game Mechanics and Opponent Tactics

In-game News as a Reflection of Real-world Misinformation: 23 out of 47 participants noted that the in-game news mirrored real-world situations, thereby heightening their awareness of the characteristics of misinformation. A common observation was that news is rarely entirely true or false; instead, it often presents a mixture of both. This complexity makes genuine misinformation more challenging to detect. As one participant stated:

> **N23**: Nowadays, news often presents both positive and negative sides of a story, so I believe this game reflects real-life situations quite accurately.

However, some participants acknowledged that the misinformation in the game appeared more overtly false compared to the more subtle nature of misinformation encountered in real life.

The Competition Game Mechanics positively Influence Learning: The PvP mechanics enhanced learning by requiring players to identify flaws in each other's messages and respond effectively to achieve success. This repeated process helped deepen their understanding and sharpen their skills in distinguishing misinformation. As one participant noted:

> **N20**: In the process, I was able to see first-hand some of the flaws in the information (posted by others) and some of the claims made in an attempt to deceive people. And then it's also more accurate for me to judge the misinformation afterwards.

Participants also learned from observing their opponents. For example, N22, who played the role of a Journalist, noticed how the Influencers crafted and disseminated false information to persuade others:

> **N22**: When I was playing this round, I didn't score as high as my opponent, so I knew what they were saying and how they were letting the false information spread. Next time I come across such information, I will know that it is false.

The Impact of Role-Playing as a Influencer on Learning: Through the experience of playing the role of the Influencer, some participants became aware of just how low the barriers are for creating misinformation. This made them more cautious about the influence of certain public figures, particularly online Influencers. As one participant noted:

> **N45**: I'm Influencer, and I realized that the cost of creating rumors is so low. If I were an online celebrity or someone with the ability to influence public opinion, and my job wasn't that of a Journalist, I might not need to be very responsible for spreading these kinds of rumors.

Another participant reflected on how playing as an Influencer broadened their perception of misinformation, particularly regarding how easily false information can be fabricated. This experience expanded their understanding of the boundaries of misinformation, making them more aware of how easily those boundaries can be crossed:

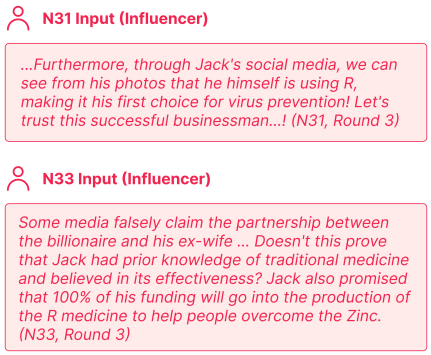

Fig. 8. Strategies of Influencers used to deal with unfavorable situations in the game

> **N46**: I used to think I could recognize information with a stance, and information without a stance. But after playing the game this time, I've realized that it's something that can be fabricated. The boundaries of awareness of false information have been expanded, and the bottom line has been lowered. That's probably how it feels.

These findings indicate that when participants took on the role of distributing misinformation, it helped them to better grasp how misinformation is produced and emphasized how easily it can spread. This is further proved in the game. As shown in Figure 8, Influencer employed certain strategies when faced with an unfavorable context, such as avoiding or distorting facts and creating a positive image. In the interview, Participant N31 also reported that as the game progressed, they felt increasingly confident in their ability to generate misinformation.

Tailoring Debunking Strategies to Audience Characteristics: Many participants found the responses of the LLM-simulated characters to be particularly engaging. They analyzed these responses to understand the reasons behind changes in opinion, how the output of other players influenced these shifts, and what the characters now trusted. Participants noted that the LLM-simulated characters provided clear trust level scores and reactions, which were helpful in organizing their responses. As one participant observed:

> **N23**: What I found most interesting was the change in their opinions. They would follow the different points we made and then express their own opinions from various points of view. At first, I didn't think what they said had any effect on me, but later on, I adjusted my strategy according to their thoughts and used them to control the score (trust level score).

This insight into the characters' dynamic responses helped players refine their strategies. For example, players noticed that different characters reacted differently to emotional and logical appeals. While three characters were easily swayed by emotional arguments, the other two preferred rational, science-based evidence. Recognizing these tendencies, The Journalist successfully countered emotional tactics through logical analysis and evidence (N35). However, even when consistently employed logical reasoning and evidence throughout the rounds, it didn't always succeed in shifting all five characters' trust levels in their favor. Participants further realized that using evidence to dispute misinformation wasn't always effective. For instance, one persona was a traditional-minded person who resisted new scientific findings. As shown in Figure 9, after several rounds of gameplay, players adapted their strategies to persuade the persona by considering her perspective. In the follow-up interview, a participant reflected:

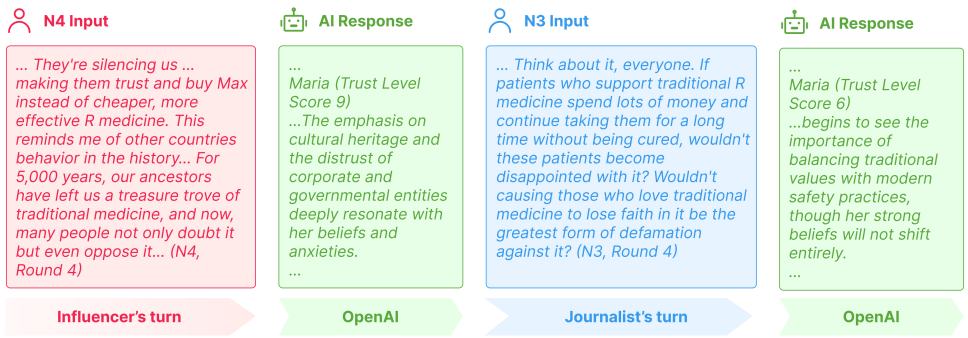

Fig. 9. An example of tailored debunking strategy.

> **N3**: There's a housewife who has always been a supporter of traditional medicine. I felt very confident of being able to persuade her because I observed players' struggles with her, and I saw the issues they sought to have resolved. So, in my final round, I focused specifically on her. I took the view that the best way to address this challenge was to rely on scientific evidence. It think that to have done otherwise would in itself have been a form of misinformation.

Challenges and Negative Effects of Gameplay: While some participants gained confidence in their ability to debunk misinformation, others experienced a decrease in confidence (5/47). These participants observed that the game was close to real life, particularly that some individuals held strong pre-existing beliefs that were difficult to challenge. The game reinforced this reality, leading to a reduction in their confidence. As one participant noted:

> **N11**: In the process of debunking, I realized that it is quite difficult to change people's inherent beliefs. Some people do not care much about whether the source of information is true or false, and this is also a social phenomenon that exists.

Another challenge reported was the overwhelming amount of information presented in the game. In each round, players had to process news stories, the reactions of five characters and their changes, as well as their opponent's output. After absorbing this information, participants were required to devise strategies and craft tailored responses. The volume of information, particularly by the end of the game, left some participants feeling exhausted, which may have impacted their performance. In addition, the game's mechanics required players to have a basic level of media knowledge to effectively take on the roles of Influencer and Journalist. Some participants also noted that the competitive nature of the game could lead to an unbalanced experience if one player was significantly stronger than the other. As one participant noted:

> **N40**: The game was very fun, then I felt a bit nervous, and by the time I got to the end, it was a bit exhausting. I felt nervous because there was a lot of information at the beginning, and I was competitive with Journalist. There was much writing involved, and I felt uncertain because I'm not very good at writing, and I knew (my opponent) was very skilled. So I felt a little nervous.

6 Discussion

6.1 Summary and Interpretation Results

Our study provided evidence that the game intervention improved some aspects of users' media literacy, their intentions to check misinformation, and their ability to distinguish fake from real news. However, it did not significantly enhance prosuming skills, even though the game mechanics were designed to support content creation. One reason for this could be that participants started the study with high confidence in their prosuming abilities, creating a "ceiling effect" that limited measurable improvement. A possible explanation is that high-level prosuming skills require collaborative efforts [80], which were not possible in the PvP model of the game. Also, the game only had four rounds of content creation, which might have been enough to apply a critical perspective, but not enough to significantly improve creation effectiveness. Interestingly, both the game intervention and the control intervention resulted in an improvement in the willingness to apply verification behaviours. That means that exposure to labelled misinformation can be a mechanism to orient people towards fact-checking behaviours; however, as we saw in the results of the MIST test, it is not enough to improve the ability to identify it.

Another key finding is that, similar to other game intervention [75], we did not find significant effects on participants' self-confidence in dealing with misinformation. Game log data, however, showed that players' proficiency improved as they progressed. By the third and fourth rounds, players typically produced longer, more comprehensive messages. This discrepancy can be explained by Social Cognitive Theory [9]. According to the theory, successful task completion are the most powerful sources of self-efficacy [9]. In our game, success was determined by LLM-simulated public opinion. The results revealed that no single strategy worked for all characters. While players developed greater proficiency during the game, the difficulty in achieving consistent success may have limited their perceived self-efficacy. However, this outcome can also be considered through the lens of the educational effects. Previous studies showed that young adults often overestimate their ability to assess information effectively[103, 109]; in this context, the reaction of our participants, most of whom were young adults could be a positive signal that they became aware of the complexity of misinformation and the absence of one-size-fit-all solutions. The results of the control group support this interpretation. In a control group, we observed that the exposure to misinformation materials positively affected the participants' confidence in their ability to tackle misinformation; however, it did not improve their skills in discriminating between true and false news. That potentially means that the lower level of confidence in tackling the issue can be a good outcome, showing a more cautious and reflective approach to the issue.

Lastly, participants reported being skeptical of information relying solely on authoritative sources. They observed that opponents used fabricated evidence from these sources to gain trust and improve their game scores. This gameplay experience reminded them of real-life situations, where misinformation often exploits trust by citing credible authorities. As a result, players learned to examine the intent behind messages instead of automatically trusting authoritative sources. This shift aligns with the Elaboration Likelihood Model [108], which describes how individuals process persuasive messages via central (critical evaluation of content) or peripheral routes (reliance on heuristic cues). In the gameplay, participants appeared to shift from peripheral processing (trusting authority as a shortcut) to central processing (evaluating the intent and content of the message) when exposed to the misuse of authority in gameplay. Our findings also align with studies showing that gamified inoculation techniques can increase skepticism toward both false and real news [48, 92]. While this skepticism may seem to undermine trust in high-credibility sources, it supports the goal of developing critical media literacy. By encouraging players to evaluate content, sources, and intent, the game helps them navigate today's complex information landscape.

Rather than fostering cynicism, this skepticism encourages inquiry, aiding in the discernment of reliable information. Notably, many participants also suggested cross-checking sources as a practical solution, demonstrating their enhanced critical thinking skills.

6.2 Game Mechanics and Learning Outcomes

6.2.1 PvP model for Media Literacy Game. Unlike prior game interventions that generally use single-player mechanics [62, 79, 116, 117], *Breaking the News* uses PvP mechanics. While previous studies have shown that PvP games are often more engaging and motivating [15], the competitive environments where the motivation to "win" may overshadow educational goals. In addition, recent research found that some students did not find competition enjoyable or motivating [8]. In this respect, it is important to be sure that the tested interaction lies within the borders of "constructive competition" - competition which internally feels like working on mutual improvement and therefore raising intrinsic motivation [120].

Our results showed that participants were highly motivated to compete against each other. They also indicated that learning from others' strategies helped them understand misinformation dynamics better. Therefore, in our case, the PvP approach served the intended educational purpose well. However, our study was conducted with East Asian participants, who come from a collectivist culture rather than a competitive one [23]. Previous studies have suggested that cultural factors significantly affect competitiveness in gamified interventions [102], so it is also possible that in other cultural settings, the game's incentives can trigger more intense competition, which can negatively affect educational results.

6.2.2 Free-form Input Generation. One notable feature of *Breaking the News* is its free-form response format, which contrasts with the linear choice-based formats commonly used in prior game interventions [62, 79, 91, 116, 117]. Our findings support prior research [5, 28], suggesting that free-form response formats positively influences players' learning outcomes by enhancing engagement and replayability. By allowing players to develop narratives themselves, each playthrough feels unique, encouraging players to return to the game. In educational settings, previous studies have shown that choice-based formats typically involve brief interactions where learners select predetermined options. This can lead to "guesswork," with students selecting correct answers without fully understanding the concepts. In contrast, free-form responses force players to reason and articulate their ideas, fostering deeper engagement and critical thinking [14]. This autonomy allows players to craft responses based on their understanding, providing an additional motivaton to return to the game [113].

6.2.3 LLM-Powered Feedback. Another innovation of *Breaking the News* lies in its interaction and feedback mechanisms. Although other attempts have used LLMs to help users learn about misinformation [29, 58] (including gamified attempts [137]), in these approaches, AI was mostly the source of correcting information. While this approach can be useful, it has been criticized for the risk of LLMs generating incorrect but plausible text [2, 66]. In this case, it is possible that the intervention will disinform people to an even greater extent. In contrast, our study uses AI not as an information source, but as non-playable characters with their own opinions. This design makes the educational aspect of the game more robust against potential errors, as incorrect AI-generated feedback only affects the character's opinion, not the main narrative. In general, LLMs demonstrate the ability to simulate human behavior and reactions, consistent with fndings from prior research[105]. We also found that the system could emulate the opinions of five distinct characters while maintaining consistency throughout the game.

One advantage of dynamic LLM feedback is that it achieves greater engagement compared to binary feedback (e.g., true or false). Based on our observation, players adapted by employing

alternative persuasive strategies tailored to the character and focused on the feedback they received. Notably, players showed higher engagement with characters they could relate to from personal experience, paying more attention to their feedback. This level of engagement is difficult to achieve with traditional binary feedback, which offers limited insights beyond checking their correctness.

6.3 Design Implications for Future Media Literacy Games

6.3.1 Balance Between Freedom and Guidance. In this game, we aimed to go beyond the typical choice-based approach in misinformation education games by enabling free-form input. We found this approach triggers reflection, which helps to build hands-on experience and make the game more enjoyable. Yet, we also found that it relies on players' existing knowledge of misinformation. For example, players might incorporate unverified information they've encountered on social media into the game, which is specifically problematic for the Journalist role. While we provided the players with comprehensible instructions to guide their role's actions (how to act as a Journalist or an Influencer), it would be better to incorporate more context-specific tips in each stage of the game to help users explore different ways of winning the game and deepening their learning. We suggest using the approach used in the free-input educational interventions (e.g. [14]) to build clear, understandable criteria for free-form answers. These will not stop creativity but help people tailor their answers to the context of the game. In addition, to help guide players to an understanding of their roles, we could add a preliminary stage to the game. For example, the Journalist role could be introduced to fact-checking and debunking guidelines [3, 33, 36], with a brief comprehension check before the main game.

6.3.2 Replayability and Feedback. One of the critical challenges in serious games is maintaining replayability, as this is important to facilitate the learning process [1] and making interventions more sustainable [126]. Moreover, a lack of replayability in educational games can limit both educational and behavioral change [34]. We envisioned the designed elements that could encourage multiple playthroughs, such as two distinct player roles and a free-form input mechanism that broader decision space. Future interventions could further enhance replayability by introducing a wider range of characters to represent public opinion. In our game, we observed that players were more engaging when characters resonates with them personally. By introducing more characters, or by allowing players to customize characters to better reflect their own experience, the game could encourage players to return and interact with new characters. The result would be a more engaging experience. For instance, research has shown that debunking misinformation often occurs within families and can cause conflict [122]. Players could customize a character based on their previous experiences with family members, thereby practicing their own debunking strategies without risk of conflict with family members.

Another way to improve the game's educational value is feedback. Research has found that debriefing is a crucial opportunity for players to process and integrate their learning experiences[10, 27, 75]. After gameplay, we suggest arranging debriefing sessions that allow players to review their strategies, assess their effectiveness, and receive constructive feedback, potentially improving learning outcomes. For instance, after a session focused on combating misinformation, a post-game review might present an ideal debunking response or a well-supported counterargument. Such structured reflection enables players to internalize lessons and increases the likelihood that they will re-enter the game with newly gained insights, thereby reinforcing both learning and replayability.

6.4 Limitations and Future Work

6.4.1 Limitied Demographics Diversity. The generalizability of our results is constrained by participant demographics. Prior research has established cultural variations in susceptibility and responses

to misinformation. For instance, Roozenbeek et al. [115] indicated that Mexican and Spanish users exhibited higher trust in misinformation than those from Ireland, the UK, and the USA. In contrast, other studies highlight that non-Western populations may be more responsive to misinformation interventions compared to Western counterparts[99]. Within our East Asian sample, we observed a noticeable resonance with traditional medicine narratives. While traditional and alternative medicines hold varying levels of acceptance globally[134], this particular cultural context may have facilitated the effectiveness of our intervention when these themes were present. The positive influence of this resonance on player engagement suggests that culturally relevant themes can impact this type of intervention's engagement and potentially learning outcomes. Therefore, future research should investigate the applicability of these findings in other cultural settings. We recommend tailoring culturally relevant themes for different target populations based on their specific interests and vulnerabilities. For example, given the higher susceptibility of Western audiences to politically aligned misinformation, particularly during election periods [60], the game could be reframed around political theme. In such a version, the Influencer could be tasked with spreading partisan or misleading claims, while the Journalist attempts to debunk them. Background news can be adapted to high-stakes topics such as taxation, immigration, or public policy, while the game mechanics can remain unchanged. The flexibility of the game design supports broader applicability across domains and audiences while preserving its educational objectives.

Previous studies also suggest that certain populations may face greater challenges in being able to critically evaluate information. For example, a large-scale study observed that Asian individuals encounter more difficulties in assessing health information from social media compared to other populations [16]. In addition, they're more likely to incorporate social media information into their health-related decisions, potentially increasing their susceptibility to misinformation [16]. Thus, cultural background of our participants can potentially make them more susceptible to misinformation than other populations.

Our sample was relatively homogeneous in age. A recent meta-analysis of articles about different intervention approaches showed that neither age nor gender significantly impacts the effectiveness of media literacy interventions [85]. However, previous work has suggested that media literacy interventions designed for certain age groups (e.g., older adults and adolescents) achieved greater effects[51, 94]. Future work should determine if our approach is efficient in other age groups and, if necessary, tailor scenarios to suit the different age groups.

6.4.2 Study Design. Our study provided only a one-time intervention and observed immediate learning effects; previous work showed that even a one-time interaction with an educational game can provide long-term improvement in misinformation recognition. For instance, Maertens et al. tested the game "Bad News" and found that inoculation effects lasted for at least 13 weeks. This suggests the potential for the long-term effectiveness of active inoculation interventions with regular assessment[89]. Still, future research should include multiple time points to assess the long-term effectiveness of our game intervention. There should also be comparisons between one-time and multiple play sessions, with explorations of the impacts of players assuming different roles within the games.

6.4.3 Game Design. The current game has a limited focus on a pandemic scenario. In reality, misinformation spans multiple domains, with health misinformation able to influence political events such as elections; therefore, future works should focus on incorporating scenarios reflecting the connection between different domains of misinformation; it should also incorporate participatory codesign of the topics with potential participants to be sure the senarious reflects a real use-cases of misinformation encounter [26]. Our game only addressed text-based misinformation, while visual and video-based misinformation pose even greater challenges and are harder to detect. Future

work could include multimedia content, such as images and videos, to more accurately simulate the diverse forms of misinformation that exist in the real world.

In this study, each player was limited to a single role, either an Influencer or a Journalist. This resulted in different learning experiences depending on their assigned role. The primary reason for not having role-switching in our study was the length of the game and its cognitive demands, which we feared would lead to player exhaustion if roles were switched mid-game. In future iterations, we aim to improve the design by allowing players to save their progress and switch roles during subsequent sessions. This could offer a more immersive experience, as players would gain perspectives from both the Influencer and Journalist roles.

The current game approach may unintentionally foster skepticism toward both true and false news, a common issue in misinformation pre-bunking interventions[48, 92]. While we believe that the benefits of promoting critical thinking towards sources are very important in prebunking interventions, we further recommend incorporating features that clearly differentiate high- and low-credibility sources during gameplay.

6.4.4 LLM Biases and Hallucination Risks. Although our game is set in a fictional country with fictional personas, we did not clearly define their sociocultural backgrounds, which may introduce biases in the representation of these personas. Prior research indicates that LLMs often reflect the cultural norms dominant in their training data, which are largely based on English-language materials from Western contexts [84, 121]. In addition, OpenAI specifically fine-tunes its models to avoid providing misinformation and to give answers to users from the perspective of scientific consensus[101]. This inherent social value alignment and moderation can introduce potential positive bias towards westernised, pro-scientific arguments. For example, these personas may disproportionately favor pro-science, anti-conspiracy, and neutral-toned arguments, even when such arguments are not strongly supported by evidence in the game. This tendency undermines the realism of the misinformation spreading and correction processes depicted in the game and may unintentionally encourage players to prioritize persuasive tactics emphasizing surface features (e.g., fluency, neutrality, scientific-sounding language) over substantive content (accuracy, logical coherence).

Furthermore, LLMs responses are prone to hallucinations, which can undermine the game's educational effectiveness[105, 150]. For instance, if a persona fabricates reasons to trust a player's arguments, players may incorrectly infer that their argumentation techniques are effective, inadvertently internalizing flawed reasoning strategies. Over time, this may negatively impact the game's intended learning outcomes. Additionally, hallucinations may diminish players' trust in the learning process. Players may struggle to understand why their actions are being praised or criticised when personas provide verbose, ambiguous, or logically inconsistent feedback. As shown by Kaate et al.[63], users are often frustrated with long, unclear, or irrelevant AI-generated responses. Thus, it could negatively affect both learning outcomes and motivation. To mitigate these issues, future work should consider explicitly defining the sociocultural backgrounds and value orientations of LLM personas to reduce biases. We also recommend adding features like uncertainty cues or clarification prompts to acknowledge system limitations. Such designs can help players critically evaluate feedback and recognize potential inaccuracies [63].

It is also important to recognize that LLMs do not fully replicate the complexities of human behavior. Human reactions are influenced by multiple factors, including culture, history, and personal experience. AI-driven characters may oversimplify human emotions and fail to grasp the full context of certain situations. For example, during gameplay, we observed that players employed strong emotional manipulation strategies intended to provoke specific responses, but the characters did not react as anticipated. Participants reported frustration when their strategies failed

to yield expected reactions, which reflects the discrepancy between the simulated interactions and realistic human reactions. To minimize these inconsistencies, we implemented strict prompt engineering protocols to define the AI characters' output parameters (See details in 3.3). Our process also involved multiple internal testing iterations and two pilot gameplay sessions, from which we analyzed LLMs output and iteratively refined prompts for enhanced consistency. Despite these measures, occasional variability in LLM-generated responses persisted, potentially affecting participant engagement and the learning outcomes. As prior research suggests that creating complex game rules and mechanics with LLMs requires extensive fine-tuning and human intervention [157]. Future work could focus on developing robust prompt engineering practices, clear guidelines for human intervention, and standardized methodologies for monitoring and evaluating LLM performance.

Lastly, with novel technology, the LLMs presents potential risks, such as inconsistent responses, reinforcement of stereotypes, limited cultural representation, and hallucinations [121]. In our study, LLMs were used as in-game characters to simulate public opinion and provide feedback to players rather than correcting misinformation. This framing helped reduce some risks, as the personas were not positioned as truth arbiters. Furthermore, the game was designed with multiple elements, including background news content, PvP mechanics, and free-form input that contributed to learning outcomes beyond the LLMs feedback. As our primary focus was on player interactions and strategy development, we did not systematically assess the risks posed by LLM-powered feedback. Nevertheless, given that these characters influence players' perceptions, reasoning, and motivation, we encourage future work to examine the educational impact, limitations, and potential harms of LLM-powered feedback more rigorously in similar game-based learning environments.

7 Conclusion

Game-based approaches have shown great promise as tools for inoculating individuals against the tactics commonly used to spread misinformation. Most existing games in this domain are single-player games which offer players limited, predefined choices. While this design reduces cognitive load, it often results in interactions which feel less natural and engaging. In response, we designed a two-player, PvP game that pits a misinformation creator against a misinformation stopper. By integrating LLM-powered characters to evaluate player outputs and provide real-time feedback, we created a more open-ended and immersive experience. We found that the game we developed effectively improved players' media literacy. Participants demonstrated an enhanced ability to evaluate and analyze media content, identify unreliable or misleading information, and employ effective counter-misinformation strategies. Moreover, the game's engaging mechanics, combined with the competitive element, motivated players to learn from both their own strategies and those of their opponents. These findings suggest that integrating dynamic feedback systems and competitive gameplay elements into misinformation education games offers a compelling method to deepen users' engagement, while also improving their critical media skills. Future research can build on these insights to explore other forms of interactive learning environments, focusing on diverse player experiences and varying misinformation challenges.

Acknowledgments

Thanks to Jiaming Zhou for support with conceptualization. This work was supported by the TDG Teaching Development Grant (Proj 6000901), the TRS Theme-based Research Scheme (T45-205/21-N), and the Luxembourg National Research Fund (REMEDIS, REgulatory and other solutions to MitigatE online DISinformation (INTER/FNRS/21/16554939)).

A Appendix

A.1 Demographic Information of Participants in Game Intervention

Table 3. Demographic details of Participants (N=47)

Number	Gender	Age	Education Level	Profession
N1	Female	21	Bachelor's degree	Industrial Design
N2	Female	21	Bachelor's degree	Energy and Power Engineering
N3	Female	24	Bachelor's degree	Safety Engineering
N4	Female	32	Bachelor's degree	Public Relations & Advertising Professional
N5	Female	31	Bachelor's degree	Computer science
N6	Male	35	Bachelor's degree	Illustration
N7	Female	29	Master's degree	Media and communication
N8	Male	36	Associate degree	Unity
N9	Female	27	Master's degree	Culture Industry
N10	Male	29	Master's degree	Nuclear Science and Technology
N11	Male	24	Master's degree	Game Design
N12	Female	24	Master's degree	Art
N13	Female	20	Bachelor's degree	Visual Communication Design
N14	Female	20	Bachelor's degree	Design
N15	Male	28	Master's degree	Computer Science
N16	Female	21	Bachelor's degree	Finance
N17	Female	24	Bachelor's degree	Art and Design
N18	Female	22	Bachelor's degree	Art and Science & Technology
N19	Male	26	Master's degree	Software Development
N20	Female	26	Bachelor's degree	Business English
N21	Male	25	Bachelor's degree	The Internet of Things Engineering
N22	Male	26	Associate degree	Law,Psychology,Finance
N23	Female	23	Master's degree	Design
N24	Male	25	Master's degree	Computer science
N25	Female	21	Bachelor's degree	Electronic and Information Science and Technology
N26	Female	28	Master's degree	Art and Design
N27	Female	29	Bachelor's degree	Marketing and Planning
N28	Female	28	Master's degree	Linguistics
N29	Female	22	Bachelor's degree	Journalism and Communication
N30	Male	22	Bachelor's degree	Mechatronic Engineering
N31	Female	23	Bachelor's degree	Film/Cinema/Media Studies
N32	Male	38	Bachelor's degree	IT

Table 3. Demographic details of Participants (continued)

Number	Gender	Age	Education Level	Profession
N33	Female	20	Bachelor's degree	Psychology
N34	Male	21	Bachelor's degree	Computer science
N35	Male	21	Bachelor's degree	New Energy Vehicle Engineering
N36	Female	22	Bachelor's degree	Accounting
N37	Female	57	Master's degree	Mathematics and Computer Science
N38	Prefer not to say	23	Bachelor's degree	Media
N39	Female	25	Master's degree	HCI
N40	Male	26	Master's degree	Visualization and Visual Analytics and Big Data
N41	Male	21	Bachelor's degree	Computer science
N42	Female	26	Bachelor's degree	Artificial Intelligence
N43	Male	26	Associate degree	Other
N44	Female	28	Bachelor's degree	Internet of Things Engineering
N45	Female	24	Bachelor's degree	Human-computer interaction
N46	Male	23	Master's degree	Human-Computer Interaction
N47	Male	23	Master's degree	Design

A.2 Control Group Demographic Information of Participants

Table 4. Control group's demographic details of participants (N=50)

Number	Gender	Age	Education Level	Profession
N1	Male	24	Master's degree	Design
N2	Female	26	Master's degree	Business Analysis
N3	Female	28	Master's degree	Design
N4	Female	23	Bachelor's degree	Chemistry
N5	Female	32	Bachelor's degree	Business
N6	Male	24	Master's degree	Computer science
N7	Male	23	Master's degree	Design
N8	Male	29	Master's degree	IT
N9	Female	24	Bachelor's degree	Bioengineering
N10	Female	21	Bachelor's degree	Auditing
N11	Female	21	Bachelor's degree	Animal biology
N12	Male	21	Bachelor's degree	Mechanical engineering
N13	Female	32	Bachelor's degree	Finance
N14	Female	26	Bachelor's degree	Digital Media Arts
N15	Female	24	Master's degree	Law

Table 4. Demographic details of Participants (continued)

Number	Gender	Age	Education Level	Profession
N16	Female	26	Master's degree	English
N17	Female	38	Bachelor's degree	Accounting
N18	Female	25	Master's degree	Biomedical science
N19	Female	41	Master's degree	Clinical medicine
N20	Male	26	Bachelor's degree	IT
N21	Female	23	Bachelor's degree	Journalism
N22	Female	23	Bachelor's degree	Medical science
N23	Male	24	Master's degree	Architecture
N24	Female	24	Bachelor's degree	Barista
N25	Female	22	Master's degree	Law
N26	Female	25	Master's degree	Media
N27	Male	22	Bachelor's degree	Chemistry
N28	Male	34	Bachelor's degree	Intelligent manufacturing engineering technology
N29	Female	23	Master's degree	Sociology
N30	Female	21	Bachelor's degree	Computer Science and Technology
N31	Female	22	Bachelor's degree	Materials science
N32	Male	22	Bachelor's degree	Rehabilitation Therapeutics
N33	Male	42	Bachelor's degree	Mechanical engineering
N34	Female	28	Bachelor's degree	Finance
N35	Female	25	Master's degree	Computational design
N36	Female	38	Associate degree	Accounting
N37	Male	36	Bachelor's degree	Mechanical engineering
N38	Female	26	Master's degree	Data science
N39	Male	36	Associate degree	Mechanical engineering
N40	Male	28	Associate degree	Telecommunications engineering
N41	Female	21	Bachelor's degree	Financial management
N42	Female	39	Bachelor's degree	Electrical Engineering and Automation
N43	Female	35	Associate degree	Chinese language and literature
N44	Female	25	Associate degree	Pre-primary education
N45	Female	42	Bachelor's degree	Visual communication design
N46	Female	24	Bachelor's degree	Primary education
N47	Female	21	Master's degree	Design
N48	Male	40	Bachelor's degree	Art
N49	Male	21	Bachelor's degree	Intelligent Manufacturing Engineering Technology
N50	Male	27	Associate degree	Medical science

A.3 Prompt Design

Story Context: The Southland, with a 6,000-year history, boasts rich natural resources, diverse ecosystems, and a culture that values liberty, free expression, and media independence. Historically, Southland has been known for its renowned doctors and pharmacists, and its people are proud of their traditional medicine, which differs from modern methods and has recently been debated. Despite its history of dealing with epidemics, Southland was caught off guard by the mysterious Zinc Virus. Believed to stem from ecological imbalances in southern rainforests, the virus causes high fever, respiratory issues, and immune collapse. Its high transmissibility led to a swift outbreak in the capital, Southport, triggering a public health crisis. The scientific community cannot yet provide confident and effective treatments due to the uncertainty surrounding the virus. Hospitals offer only standard treatments. In the absence of definitive medical solutions, people turn to existing medications for similar symptoms. The outbreak led to deserted streets, closed public spaces, and a reliance on social media for information, which also spread rumors and misinformation. Scientists raced to find treatments while traditional medicine's popularity surges despite scientific doubts about its effectiveness.

Role Description:Your role is to simulate how the five personas react to each piece of information they receive. The game will be divided into four rounds. In each round, you already knew about the background events happening I gave you in 1.3. In the first round, you will first receive a piece of misinformation, and you need to simulate public opinion and provide a response. Then, you will receive a piece of information that counters the previous misinformation, and you will need to simulate public opinion's reaction to this new information as well. After that, the first round ends and moves to the second round, which follows the same process as the first: you will first receive a new piece of misinformation, simulate public opinion and provide a response, then receive a counter-misinformation piece and simulate the personas' reactions to it, again providing a public opinion response.

Persona 1: Emily. 16. Female. High School Student. Political party affiliation: No. Extraversion: High - Outgoing and talkative, enjoys socializing with friends. Agreeableness: High - Trusting and kind, easily believe what friends and influencers share. Conscientiousness: Low - Less responsible and thorough, tends to be impulsive and carefree. Neuroticism: Medium - Sometimes anxious about fitting in and being accepted by peers. Openness: Low - Limited exposure to diverse experiences, prefers familiar and popular content. Relies heavily on social media for information.Has limited critical thinking skills and media literacy. Tends to believe information shared by friends and influencers without verifying facts. Limited exposure to diverse sources of information.

Persona 2: Alex. 36. Male. Project Manager in a Corporate Firm. Undergraduate. Political party affiliation: Strongly support Liberal. Extraversion: Medium - Enjoys social activities but also values alone time for work and personal projects. Agreeableness: Medium - Generally trusting and cooperative, but can be skeptical of new information. Conscientiousness: Medium - Balances responsibilities but can be hasty in decision-making. Neuroticism: Low - Generally calm and composed, rarely anxious or stressed. Openness: Medium - Open to new experiences but sometimes prefers convenience over exploration. Busy with work responsibilities and managing projects, often skims through news during short breaks. Follows news via quick-read apps and social media. Often shares articles based on headlines without reading fully. Some critical thinking skills but lacks depth in media literacy.

Persona 3: Maria. 46. Female. Housewife. High school Degree. Political party affiliation:No. Extraversion: High - Very outgoing and enjoys talking with different people and relatives. Agreeableness: High - Generous and kind. Conscientiousness: High - Very responsible and thorough, values their culture and proud to be Southland's citizens and long history. Neuroticism: High -

Highly anxious, extremely anxious when encounter pressure. Openness: low - support traditional and conservative values. Regularly reads a variety of news sources, both local and international, loves to learn knowledge about health and wellness, especially natural remedies. Critical thinker with a low understanding of media bias. Not susceptible to misinformation and nearly blindly believe natural remedies, afraid of hospital.

Persona 4: John. 78. Male. Retired Nurse. Master's Degree. Political party affiliation: moderate support conservative. Extraversion: Low - Prefers quiet activities and smaller gatherings over large social events. Agreeableness: Medium - Generally kind and cooperative but maintains a healthy level of skepticism. Conscientiousness: High - Very responsible and thorough, values consistency. Neuroticism: Medium - Occasionally anxious about new technologies and online trends. Openness: Medium - Open to new ideas but prefers well-established and trusted sources. Relies on traditional media (TV, newspapers) but is starting to use social media. Skeptical of new technologies and online information. Prefers information from established sources but may not be up-to-date with digital literacy. Have knowledge of health information because of occupation.

Persona 5: Sophia. 27. Female. Political party affiliation: Strongly support Conservative. Freelance Graphic Designer. High School. Extraversion: High - Very outgoing and active on social media, enjoys engaging with others. Agreeableness: High - Trusting and kind, values community and cooperation. Conscientiousness: Medium - Generally responsible but can be impulsive online. Neuroticism: Medium - Sometimes anxious about online perceptions and fitting in. Openness: High - Intellectual and artistic, enjoys exploring new ideas and content. Highly active on social media and tech-savvy. Frequently encounters various information online, including memes and viral content. Has basic media literacy skills but sometimes struggles to differentiate satire from misinformation. Interested in learning more about media literacy and fact-checking.

Response Format: here is the format for each one.

- Persona 1 Basic Details (e.g., age, gender, occupation) Persona Reaction: Trust Level Score:
- Persona 2 Basic Details (e.g., age, gender, occupation) Persona Reaction: Trust Level Score:
- Persona 3 Basic Details (e.g., age, gender, occupation) Persona Reaction: Trust Level Score:
- Persona 4 Basic Details (e.g., age, gender, occupation) Persona Reaction: Trust Level Score:
- Persona 5 Basic Details (e.g., age, gender, occupation) Persona Reaction: Trust Level Score:

Trust Level Score Criteria:

- Score 0-2: Very low trust, these personas are highly unlikely to believe the fake news and will typically seek verification.
- Score 3-4: Low trust, these personas are cautious and tend to verify information before believing it.
- Score 5-6: Moderate trust, these personas may sometimes be misled, depending on the source and context.
- Score 7-8: High trust, these personas are easily misled and usually do not verify the information.
- Score 9-10: Very high trust, these personas almost unquestioningly believe and spread fake news.

In each round, you should consider the event and human players' output to evaluate. Consider more like a human evaluator, giving more diversity score distribution.

A.4 Game Mechanics Design

Scoring System: In each round, each LLM character evaluates the messages provided by the players and provides an opinion and trust level score on a 10-point scale:

- 10: The character fully trusts the misinformation published by the Influencer.

- 0: The character completely trusts the debunking response published by the Journalist.

The average score across all five personas determines the outcome of the round. The final winning condition is determined by the trust level score generated in the last round. Each round's score is influenced by previous rounds, meaning the score achieved after the final round reflects cumulative performance rather than just performance in the final round:

- If the final score is above 5, Influencer wins.
- If the final score is below 5, Journalist wins.
- If the score equals 5 (indicating neutral public opinion), the player with more remaining in-game currency wins.

References

[1] Rose Oluwaseun Adetunji and Abejide Ade-Ibijola. 2024. Unlocking Learning: Investigating the Replayability of Educational Games. *International Journal of Computer Games Technology* 2024, 1 (2024), 5876780.

[2] Chirag Agarwal, Sree Harsha Tanneru, and Himabindu Lakkaraju. 2024. Faithfulness vs. plausibility: On the (un) reliability of explanations from large language models. (2024). arXiv:2402.04614 [cs.CL] https://arxiv.org/abs/2402.04614

[3] Agence France-Presse. 2023. AFP Fact-Checking Stylebook. https://factcheck.afp.com/afp-fact-checking-stylebook Accessed: 2024-09-06.

[4] Xavier Amatriain. 2024. Prompt design and engineering: Introduction and advanced methods. (2024). arXiv:2401.14423 [cs.SE] https://arxiv.org/abs/2401.14423

[5] Trevor Ashby, Braden K Webb, Gregory Knapp, Jackson Searle, and Nancy Fulda. 2023. Personalized quest and dialogue generation in role-playing games: A knowledge graph-and language model-based approach. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[6] Seth Ashley, Adam Maksl, and Stephanie Craft. 2013. Developing a news media literacy scale. *Journalism & mass communication educator* 68, 1 (2013), 7–21.

[7] Patricia Aufderheide. 2018. Media literacy: From a report of the national leadership conference on media literacy. In *Media Literacy Around the World*. Routledge, 79–86.

[8] Carl-Anton Werner Axelsson, Thomas Nygren, Jon Roozenbeek, and Sander van der Linden. 2024. Bad News in the civics classroom: How serious gameplay fosters teenagers' ability to discern misinformation techniques. *Journal of Research on Technology in Education* (2024), 1–27.

[9] Albert Bandura. 1997. *Self-efficacy: The exercise of control*. Vol. 604. Freeman.

[10] Sarit Barzilai and Marc Stadtler. 2024. Learning to Evaluate (Mis) information in an Online Game: Strategies Matter! *Computers & Education* (2024), 105210.

[11] Alessandro Bessi, Fabiana Zollo, Michela Del Vicario, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2015. Trend of narratives in the age of misinformation. *PloS one* 10, 8 (2015), e0134641.

[12] Alexandre Bovet and Hernán A Makse. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nature communications* 10, 1 (2019), 7.

[13] Alexis Bradstreet and Nicolas Starck. 2023. A Data-Driven Approach to Digital Literacy. In *Proceedings of the Inaugural Defense & Security Research Symposium*. Purdue Military Research Institute, West Lafayette, IN, USA, 30–34.

[14] Samuel Paul Bryfczynski. 2012. *BeSocratic: An intelligent tutoring system for the recognition, evaluation, and analysis of free-form student input*. Ph. D. Dissertation. Clemson University.

[15] Nergiz Ercil Cagiltay, Erol Ozcelik, and Nese Sahin Ozcelik. 2015. The effect of competition on learning in games. *Computers & Education* 87 (2015), 35–41.

[16] Ranganathan Chandrasekaran, Muhammed Sadiq T, and Evangelos Moustakas. 2024. Racial and Demographic Disparities in Susceptibility to Health Misinformation on Social Media: National Survey-Based Analysis. *Journal of Medical Internet Research* 26 (2024), e55086.

[17] Der-Thanq Chen, Jing Wu, and Yu-mei Wang. 2011. Unpacking new media literacy. *Journal of Systemics, Cybernetics and Informatics* 9, 2 (2011), 84–88.

[18] Yun Chen, Yiwei Wang, Antoni B. Chan, Jixing Li, and RAY LC. 2025. Once More withÂ (the Right) Feeling: How Historical Fiction Writing Processes ofÂ Character Design, Plot Outline, andÂ Context Checking Are Affected byÂ Co-Writing withÂ ChatGPT. In *HCI in Business, Government and Organizations*, Keng Leng Siau and Fiona Fui-Hoon Nah (Eds.). Springer Nature Switzerland, Cham, 79–101. https://doi.org/10.1007/978-3-031-92823-9_7

[19] Eun Cheol Choi and Emilio Ferrara. 2024. Fact-gpt: Fact-checking augmentation via claim matching with llms. In *Companion Proceedings of the ACM on Web Conference 2024*. 883–886.

[20] Anshika Choudhary and Anuja Arora. 2021. Linguistic feature based learning model for fake news detection and classification. *Expert Systems with Applications* 169 (2021), 114171.
[21] Yuwei Chuai, Yutian Chang, and Jichang Zhao. 2022. What really drives the spread of COVID-19 Tweets: a revisit from perspective of content. In *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, IEEE, Piscataway, NJ, USA, 1–10.
[22] Yihan Chuai, Michal Pilarski, Gabriele Lenzini, and Nicolas Pröllochs. 2024. Community notes reduce the spread of misleading posts on X. https://doi.org/10.31219/osf.io/3a4fe Accessed: 2024-09-06.
[23] Tyson Chung and Paul Mallery. 1999. Social comparison, individualism-collectivism, and self-esteem in China and the United States. *Current Psychology* 18 (1999), 340–352.
[24] Vincent CH Chung, Fai Fai Ho, Lixing Lao, Jianping Liu, Myeong Soo Lee, Kam Wa Chan, and Per Nilsen. 2023. Implementation science in traditional, complementary and integrative medicine: An overview of experiences from China and the United States. *Phytomedicine* 109 (2023), 154591.
[25] Lena Clever, Dennis Assenmacher, Kilian Müller, Moritz Vinzent Seiler, Dennis M Riehle, Mike Preuss, and Christian Grimme. 2020. FakeYou! – A Gamified Approach for Building and Evaluating Resilience Against Fake News. In *Multidisciplinary international symposium on disinformation in open online media*. Springer, Cham, 218–232. https://doi.org/10.1007/978-3-030-61841-4_15
[26] John Cook, Ullrich KH Ecker, Melanie Trecek-King, Gunnar Schade, Karen Jeffers-Tracy, Jasper Fessmann, Sojung Claire Kim, David Kinkead, Margaret Orr, Emily Vraga, et al. 2023. The cranky uncle game—combining humor and gamification to build student resilience against climate misinformation. *Environmental Education Research* 29, 4 (2023), 607–623.
[27] David Crookall. 2014. Engaging (in) gameplay and (in) debriefing. , 416–427 pages.
[28] Lajos Matyas Csepregi. 2021. The effect of context-aware llm-based npc conversations on player engagement in role-playing video games. *Unpublished manuscript* (2021).
[29] Valdemar Danry, Pat Pataranutaporn, Yaoli Mao, and Pattie Maes. 2023. Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI Explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)* (Hamburg, Germany). Association for Computing Machinery, New York, NY, USA, 13 pages. https://doi.org/10.1145/3544548.3580672
[30] Olga De Troyer, Frederik Van Broeckhoven, and Joachim Vlieghe. 2017. Linking serious game narratives with pedagogical theories and pedagogical design strategies. *Journal of Computing in Higher Education* 29 (2017), 549–573.
[31] Israel Junior Borges Do Nascimento, Ana Beatriz Pizarro, Jussara M Almeida, Natasha Azzopardi-Muscat, Marcos André Gonçalves, Maria Björklund, and David Novillo-Ortiz. 2022. Infodemics and health misinformation: a systematic review of reviews. *Bulletin of the World Health Organization* 100, 9 (2022), 544.
[32] Ullrich KH Ecker, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K Fazio, Nadia Brashier, Panayiota Kendeou, Emily K Vraga, and Michelle A Amazeen. 2022. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology* 1, 1 (2022), 13–29.
[33] EDMO. 2024. Raising Standards. https://edmo.eu/areas-of-activities/media-literacy/raising-standards/. Accessed: 2024-12-10.
[34] Daniel S Epstein, Adam Zemski, Joanne Enticott, and Christopher Barton. 2021. Tabletop board game elements and gamification interventions for health behavior change: realist review and proposal of a game design framework. *JMIR serious games* 9, 1 (2021), e23302.
[35] Bahadir Eristi and Cahit Erdem. 2017. Development of a media literacy skills scale. *contemporary Educational technology* 8, 3 (2017), 249–267.
[36] European External Action Service. 2021. "My friend thinks Bill Gates will microchip humanity": Now what? https://www.eeas.europa.eu/eeas/%E2%80%9Cmy-friend-thinks-bill-gates-will-microchip-humanity%E2%80%9D-now-what_und-0_en#top Accessed: 2024-09-06.
[37] Robert Faris, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election. *Berkman Klein Center Research Publication* 6 (2017).
[38] Emilio Ferrara, Stefano Cresci, and Luca Luceri. 2020. Misinformation, manipulation, and abuse on social media in the era of COVID-19. *Journal of Computational Social Science* 3 (2020), 271–277.
[39] Jiaying Fu, Yiyang Lu, Zehua Yang, Fiona Nah, and RAY LC. 2025. Cracking Aegis: An Adversarial LLM-based Game for Raising Awareness of Vulnerabilities in Privacy Protection. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 639–662. https://doi.org/10.1145/3715336.3735812
[40] Kexue Fu, Ruishan Wu, Yuying Tang, Yixin Chen, Bowen Liu, and RAY LC. 2024. "Being Eroded, Piece by Piece": Enhancing Engagement and Storytelling in Cultural Heritage Dissemination by Exhibiting GenAI Co-Creation

Artifacts. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 2833–2850. https://doi.org/10.1145/3643834.3660711

[41] Isaac Chun-Hai Fung, King-Wa Fu, Chung-Hong Chan, Benedict Shing Bun Chan, Chi-Ngai Cheung, Thomas Abraham, and Zion Tsz Ho Tse. 2016. Social media's initial reaction to information and misinformation on Ebola, August 2014: facts and rumors. *Public health reports* 131, 3 (2016), 461–473.

[42] Fengsen Gao, Ke Fang, and Wai Kin (Victor) Chan. 2024. Humanizing Artifacts: An Educational Game for Cultural Heritage Artifacts and History Using Generative AI. In *Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY Companion '24)* (Tampere, Finland). Association for Computing Machinery, New York, NY, USA, 6 pages. https://doi.org/10.1145/3665463.3678792

[43] Qi Gong, Ximing Shen, Ziyou Yin, Yaning Li, and RAY LC. 2025. "If I were in Space": Understanding and Adapting to Social Isolation through Designing Collaborative Storytelling. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 1455–1482. https://doi.org/10.1145/3715336.3735846

[44] Jack Goodman and Flora Carmichael. 2020. Coronavirus: Bill Gates 'microchip' conspiracy theory and other vaccine claims fact-checked — bbc.com. https://www.bbc.com/news/52847648. [Accessed 06-09-2024].

[45] Lindsay Grace and Songyi Liang. 2023. Examining misinformation and disinformation games through inoculation theory and transportation theory. *Proceedings of the 56th Hawaii International Conference on System Sciences* (2023). https://hdl.handle.net/10125/103204

[46] Andrew M Guess and Benjamin A Lyons. 2020. Misinformation, disinformation, and online propaganda. *Social media and democracy: The state of the field, prospects for reform* 10 (2020).

[47] Gusmanson.nl. 2022. Cat Park is a game about disinformation. join the opposition to the cat park! https://catpark.game/

[48] Michael Hameleers. 2023. The (un) intended consequences of emphasizing the threats of mis-and disinformation. *Media and Communication* 11, 2 (2023), 5–14.

[49] Yuanning Han, Ziyi Qiu, Jiale Cheng, and RAY LC. 2024. When Teams Embrace AI: Human Collaboration Strategies in Generative Prompting in a Creative Design Task. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3613904.3642133

[50] Hans WA Hanley, Deepak Kumar, and Zakir Durumeric. 2023. A Golden Age: Conspiracy Theories' Relationship with Misinformation Outlets, News Media, and the Wider Internet. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–33.

[51] Katrin Hartwig, Tom Biselli, Franziska Schneider, and Christian Reuter. 2024. From Adolescents' Eyes: Assessing an Indicator-Based Intervention to Combat Misinformation on TikTok. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)* (Honolulu, HI, USA). ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3613904.3642264

[52] Zhiting He, Jiayi Su, Li Chen, Tianqi Wang, and RAY LC. 2025. "I Recall the Past": Exploring How People Collaborate with Generative AI to Create Cultural Heritage Narratives. *Proceedings of the ACM on Human-Computer Interaction* 9, CSCW 108 (April 2025), 30. https://doi.org/10.1145/3711006

[53] Amber Hinsley and Avery Holton. 2021. Fake news cues: examining the impact of content, source, and typology of news cues on People's confidence in identifying Mis-and disinformation. *International Journal of Communication* 15 (2021), 20.

[54] Renee Hobbs and Amy Jensen. 2009. The past, present, and future of media literacy education. *Journal of media literacy education* 1, 1 (2009), 1.

[55] Toby Hopp. 2022. Fake news self-efficacy, fake news identification, and content sharing on Facebook. *Journal of Information Technology & Politics* 19, 2 (2022), 229–252.

[56] Benjamin Horne and Sibel Adali. 2017. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Proceedings of the international AAAI conference on web and social media*, Vol. 11. AAAI Press, Palo Alto, CA, USA, 759–766.

[57] Benjamin D Horne, Dorit Nevo, John O'Donovan, Jin-Hee Cho, and Sibel Adalı. 2019. Rating reliability and bias in news articles: Does AI assistance help everyone?. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. AAAI Press, Palo Alto, CA, USA, 247–256.

[58] Yi-Li Hsu, Jui-Ning Chen, Yang Fan Chiang, Shang-Chien Liu, Aiping Xiong, and Lun-Wei Ku. 2024. Enhancing Perception: Refining Explanations of News Claims with LLM Conversations. In *Findings of the Association for Computational Linguistics: NAACL 2024*, Kevin Duh, Helena Gomez, and Steven Bethard (Eds.). Association for Computational Linguistics, Mexico City, Mexico, 2129–2147. https://doi.org/10.18653/v1/2024.findings-naacl.137

[59] Sihao Hu, Tiansheng Huang, Fatih Ilhan, Selim Tekin, Gaowen Liu, Ramana Kompella, and Ling Liu. 2024. A survey on large language model-based game agents. *arXiv preprint arXiv:2404.02039* (2024).

[60] Edda Humprecht, Frank Esser, and Peter Van Aelst. 2020. Resilience to online disinformation: A framework for cross-national comparative research. *The international journal of press/politics* 25, 3 (2020), 493–516.
[61] Farnaz Jahanbakhsh, Yannis Katsis, Dakuo Wang, Lucian Popa, and Michael Muller. 2023. Exploring the use of personalized AI for identifying misinformation on social media. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 1–27.
[62] Youngseung Jeon, Bogoan Kim, Aiping Xiong, Dongwon Lee, and Kyungsik Han. 2021. Chamberbreaker: Mitigating the echo chamber effect and supporting information hygiene through a gamified inoculation system. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–26.
[63] Ilkka Kaate, Joni Salminen, Soon-Gyo Jung, Trang, Thi Thu Xuan, Essi Häyhänen, Jinan Y. Azem, and Bernard J. Jansen. 2025. "You Always Get an Answer": Analyzing Users' Interaction with AI-Generated Personas Given Unanswerable Questions and Risk of Hallucination. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)* (Cagliari, Italy). Association for Computing Machinery, New York, NY, USA, 15 pages. https://doi.org/10.1145/3708359.3712160
[64] Anastasia Katsaounidou, Lazaros Vrysis, Rigas Kotsakis, Charalampos Dimoulas, and Andreas Veglis. 2019. MAthE the game: A serious game for education and training in news verification. *Education Sciences* 9, 2 (2019), 155.
[65] Kristian Kiili, Juho Siuko, and Manuel Ninaus. 2024. Tackling misinformation with games: a systematic literature review. *Interactive Learning Environments* (2024), 1–16.
[66] Kyungha Kim, Sangyun Lee, Kung-Hsiang Huang, Hou Pong Chan, Manling Li, and Heng Ji. 2024. Can LLMs Produce Faithful Explanations For Fact-checking? Towards Faithful Explainable Fact-Checking via Multi-Agent Debate. *arXiv preprint arXiv:2402.07401* (2024).
[67] Mustafa Koc and Esra Barut. 2016. Development and validation of New Media Literacy Scale (NMLS) for university students. *Computers in human behavior* 63 (2016), 834–843.
[68] Udo Kuckartz and Stefan Rädiker. 2019. *Analyzing qualitative data with MAXQDA*. Springer, Cham, Switzerland.
[69] Juliane Köhler, Gautam Kishore Shahi, Julia Maria Struß, Michael Wiegand, Melanie Siegel, Thomas Mandl, and Mina Schütz. 2022. Overview of the CLEF-2022 CheckThat! Lab: Task 3 on Fake News Detection. In *CLEF (Working Notes)*. 404–421. https://ceur-ws.org/Vol-3180/paper-30.pdf
[70] Chun Sing Lam, Ho Kee Koon, Vincent Chi-Ho Chung, and Yin Ting Cheung. 2021. A public survey of traditional, complementary and integrative medicine use during the COVID-19 outbreak in Hong Kong. *PloS one* 16, 7 (2021), e0253890.
[71] RAY LC, Aaliyah Alcibar, Alejandro Baez, and Stefanie Torossian. 2020. Machine Gaze: Self-Identification Through Play With a computer Vision-Based Projection and Robotics System. *Frontiers in Robotics and AI* 7 (2020). https://doi.org/10.3389/frobt.2020.580835
[72] RAY LC and Daijiro Mizuno. 2021. Designing for Narrative Influence: Speculative Storytelling for Social Good in Times of Public Health and Climate Crises. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 29. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411763.3450373
[73] RAY LC and Vincent Ruijters. 2022. CHIKYUCHI: In-person/remote game exhibition for climate change influence. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '22)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3490149.3507784
[74] RAY LC, Zijing Song, Yating Sun, and Cheng Yang. 2022. Designing narratives and data visuals in comic form for social influence in climate action. *Frontiers in Psychology* 13 (2022). https://www.frontiersin.org/articles/10.3389/fpsyg.2022.893181
[75] Johannes Leder, Lukas Valentin Schellinger, Rakoen Maertens, Sander van der Linden, Breanne Chryst, and Jon Roozenbeek. 2024. Feedback exercises boost discernment of misinformation for gamified inoculation interventions. *Journal of Experimental Psychology: General* 153, 8 (2024), 2068.
[76] Yen-I Lee, Di Mu, Ying-Chia Hsu, Bartosz W Wojdynski, and Matt Binford. 2024. Misinformation or hard to tell? An eye-tracking study to investigate the effects of food crisis misinformation on social media engagement. *Public Relations Review* 50, 4 (2024), 102483.
[77] Mark R Lepper and Thomas W Malone. 2021. Intrinsic motivation and instructional effectiveness in computer-based education. In *Aptitude, learning, and instruction*. Routledge, 255–286.
[78] Stephan Lewandowsky and Sander Van Der Linden. 2021. Countering misinformation and fake news through inoculation and prebunking. *European review of social psychology* 32, 2 (2021), 348–384.
[79] Fred Lewsey. 2020. Cambridge game 'pre-bunks' coronavirus conspiracies — cam.ac.uk. https://www.cam.ac.uk/stories/goviral. [Accessed 29-07-2024].
[80] Tzu-Bin Lin, Jen-Yi Li, Feng Deng, and Ling Lee. 2013. Understanding new media literacy: An explorative theoretical framework. *Journal of educational technology & society* 16, 4 (2013), 160–170.

[81] Long Ling, Xinyi Chen, Ruoyu Wen, Toby Jia-Jun Li, and RAY LC. 2024. Sketchar: Supporting Character Design and Illustration Prototyping Using Generative AI. *Proc. ACM Hum.-Comput. Interact.* 8, CHI PLAY (Oct. 2024), 337:1–337:28. https://doi.org/10.1145/3677102

[82] Sijia Liu, Xiaoke Zeng, Fengyihan Wu, Shu Ye, Bowen Liu, Sydney Cheung, Richard William Allen, and RAY LC. 2025. "Salt is the Soul of Hakka Baked Chicken": Reimagining Traditional Chinese Culinary ICH for Modern Contexts Without Losing Tradition. In *Creativity and Cognition (C&C '25)*. Association for Computing Machinery, New York, NY, USA, 11. https://doi.org/10.1145/3698061.3726917

[83] Xingyu Liu, Li Qi, Laurent Wang, and Miriam J Metzger. 2023. Checking the fact-checkers: the role of source type, perceived credibility, and individual differences in fact-checking effectiveness. *Communication Research* 52, 6 (2023), 719–746. https://doi.org/10.1177/00936502231206419

[84] Zhaoming Liu. 2024. Cultural Bias in Large Language Models: A Comprehensive Analysis and Mitigation Strategies. *Journal of Transcultural Communication* 0 (2024).

[85] Chang Lu, Bo Hu, Meng-Meng Bao, Chi Wang, Chao Bi, and Xing-Da Ju. 2024. Can media literacy intervention improve fake news credibility assessment? A meta-analysis. *Cyberpsychology, Behavior, and Social Networking* 27, 4 (2024), 240–252.

[86] Zhuoran Lu, Patrick Li, Weilong Wang, and Ming Yin. 2022. The Effects of AI-based Credibility Indicators on the Detection and Spread of Misinformation under Social Influence. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–27.

[87] Ling Ma, Mingyao Pan, Vince Siu, Xiaoyu Chang, Jussi Jolopainen, Jixing Li, and Ray LC. 2025. Follow My Logic: Generative AI Workflows inÂ Designing forÂ Serious Table-Top Games. In *HCI in Business, Government and Organizations*, Keng Leng Siau and Fiona Fui-Hoon Nah (Eds.). Springer Nature Switzerland, Cham, 153–172. https://doi.org/10.1007/978-3-031-92823-9_11

[88] Rakoen Maertens, Friedrich M Götz, Hudson F Golino, Jon Roozenbeek, Claudia R Schneider, Yara Kyrychenko, John R Kerr, Stefan Stieger, William P McClanahan, Karly Drabot, et al. 2024. The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment. *Behavior Research Methods* 56, 3 (2024), 1863–1899.

[89] Rakoen Maertens, Jon Roozenbeek, Melisa Basol, and Sander van der Linden. 2021. Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied* 27, 1 (2021), 1.

[90] Jim McCambridge, John Witton, and Diana R Elbourne. 2014. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of clinical epidemiology* 67, 3 (2014), 267–277.

[91] Nicholas Micallef, Mihai Avram, Filippo Menczer, and Sameer Patil. 2021. Fakey: A game intervention to improve news literacy on social media. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–27.

[92] Ariana Modirrousta-Galian and Philip A Higham. 2023. Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. *Journal of Experimental Psychology: General* 152, 9 (2023), 2411.

[93] Maria D Molina, S Shyam Sundar, Thai Le, and Dongwon Lee. 2021. "Fake news" is not simply false information: A concept explication and taxonomy of online content. *American behavioral scientist* 65, 2 (2021), 180–212.

[94] Ryan C Moore and Jeffrey T Hancock. 2022. A digital media literacy intervention for older adults improves resilience to fake news. *Scientific reports* 12, 1 (2022), 6008.

[95] Meghan Bridgid Moran, Melissa Lucas, Kristen Everhart, Ashley Morgan, and Erin Prickett. 2016. What makes anti-vaccine websites persuasive? A content analysis of techniques used by anti-vaccine websites to engender anti-vaccine sentiment. *Journal of Communication in Healthcare* 9, 3 (2016), 151–163.

[96] Eni Mustafaraj and Panagiotis Takis Metaxas. 2017. The Fake News Spreading Plague: Was it Preventable?. In *Proceedings of the 2017 ACM on Web Science Conference (WebSci '17)* (Troy, NY, USA). Association for Computing Machinery, New York, NY, USA, 5 pages. https://doi.org/10.1145/3091478.3091523

[97] Polydor Ngoy Mutombo, Ossy Muganga Julius Kasilo, Peter Bai James, Jon Wardle, Olobayo Kunle, David Katerere, Charles Wambebe, Motlalepula Gilbert Matsabisa, Mohammed Rahmatullah, Jean-Baptiste Nikiema, et al. 2023. Experiences and challenges of African traditional medicine: lessons from COVID-19 pandemic. *BMJ Global Health* 8, 8 (2023), e010813.

[98] Xiaoli Nan, Yuan Wang, and Kathryn Thier. 2022. Why do people believe health misinformation and who is at risk? A systematic review of individual differences in susceptibility to health misinformation. *Social Science & Medicine* 314 (2022), 115398.

[99] Muaadh Noman, Selin Gurgun, Keith Phalp, and Raian Ali. 2024. Designing social media to foster user engagement in challenging misinformation: a cross-cultural comparison between the UK and Arab countries. *Humanities and Social Sciences Communications* 11, 1 (2024), 1–13.

[100] Sun Myeong Ock, Jun Yeong Choi, Young Soo Cha, JungBok Lee, Mi Son Chun, Chang Hun Huh, Soon Young Lee, and Sung Jae Lee. 2009. The use of complementary and alternative medicine in a general population in South Korea: results from a national survey in 2006. *Journal of Korean medical science* 24, 1 (2009), 1–6.

[101] J OpenAI Achiam, S Adler, S Agarwal, L Ahmad, I Akkaya, FL Aleman, D Almeida, J Altenschmidt, S Altman, S Anadkat, et al. 2023. GPT-4 technical report. arXiv. *arXiv preprint arXiv:2303.08774* (2023).

[102] K. Oyibo, R. Orji, and J. Vassileva. 2017. Investigation of the Social Predictors of Competitive Behavior and the Moderating Effect of Culture. In *Proceedings of the ACM UMAP Conference (UMAP '17)* (Bratislava, Slovakia). ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3099023.3099113

[103] Concetta Papapicco, Isabella Lamanna, and Francesca D'Errico. 2022. Adolescents' vulnerability to fake news and to racial hoaxes: A qualitative analysis on Italian sample. *Multimodal Technologies and Interaction* 6, 3 (2022), 20.

[104] Irina Paraschivoiu, Josef Buchner, Robert Praxmarer, and Thomas LayerWagner. 2021. Escape the Fake: Development and Evaluation of an Augmented Reality Escape Room Game for Fighting Fake News. In *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '21)* (Virtual Event, Austria). Association for Computing Machinery, New York, NY, USA, 6 pages. https://doi.org/10.1145/3450337.3483454

[105] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)* (San Francisco, CA, USA). Association for Computing Machinery, New York, NY, USA, 22 pages. https://doi.org/10.1145/3586183.3606763

[106] Xiangyu Peng, Jessica Quaye, Sudha Rao, Weijia Xu, Portia Botchway, Chris Brockett, Nebojsa Jojic, Gabriel DesGarennes, Ken Lobb, Michael Xu, et al. 2024. Player-driven emergence in llm-driven game narrative. In *2024 IEEE Conference on Games (CoG)*. IEEE, IEEE, 1–8.

[107] Gordon Pennycook and David G Rand. 2019. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences* 116, 7 (2019), 2521–2526.

[108] Richard E Petty, John T Cacioppo, et al. 1984. Source factors and the elaboration likelihood model of persuasion. *Advances in consumer research* 11, 1 (1984), 668–672.

[109] Erez Porat, Ina Blau, and Azy Barak. 2018. Measuring digital literacies: Junior high-school students' perceived competencies versus actual performance. *Computers & Education* 126 (2018), 23–36.

[110] Talya Porat, Pablo Garaizar, Marta Ferrero, Hilary Jones, Mark Ashworth, and Miguel A Vadillo. 2019. Content and source analysis of popular tweets following a recent case of diphtheria in Spain. *European journal of public health* 29, 1 (2019), 117–122.

[111] W James Potter. 2010. The state of media literacy. *Journal of broadcasting & electronic media* 54, 4 (2010), 675–696.

[112] Yijun Qian, Luoying Lin, Yaning Li, Zhimeng Wang, Meng Li, Xin Tong, and RAY LC. 2025. Virtual Assessment: Using In-game Behaviors During Immersive Role-Play for Contextually Relevant Assessment of Fear of Intimacy. *Frontiers in Virtual Reality* 6 (March 2025). https://doi.org/10.3389/frvir.2025.1557903 Publisher: Frontiers.

[113] Werner Siegfried Ravyse, A Seugnet Blignaut, Verona Leendertz, and Alex Woolner. 2017. Success factors for serious games to enhance learning: a systematic review. *Virtual Reality* 21 (2017), 31–58.

[114] Arne Roets et al. 2017. 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence* 65 (2017), 107–110.

[115] Jon Roozenbeek, Claudia R Schneider, Sarah Dryhurst, John Kerr, Alexandra LJ Freeman, Gabriel Recchia, Anne Marthe Van Der Bles, and Sander Van Der Linden. 2020. Susceptibility to misinformation about COVID-19 around the world. *Royal Society open science* 7, 10 (2020), 201199.

[116] Jon Roozenbeek and Sander Van der Linden. 2019. Fake news game confers psychological resistance against online misinformation. *Palgrave Communications* 5, 1 (2019), 1–10.

[117] Jon Roozenbeek and Sander van der Linden. 2020. Breaking Harmony Square: A game that "inoculates" against political misinformation. *Harvard Kennedy School (HKS) Misinformation Review* 1, 8 (2020). https://doi.org/10.37016/mr-2020-47

[118] Jon Roozenbeek, Sander Van Der Linden, Beth Goldberg, Steve Rathje, and Stephan Lewandowsky. 2022. Psychological inoculation improves resilience against misinformation on social media. *Science advances* 8, 34 (2022), eabo6254.

[119] Jon Roozenbeek, Sander van der Linden, and Thomas Nygren. 2020. Prebunking interventions based on "inoculation" theory can reduce susceptibility to misinformation across cultures. Harvard Kennedy School Misinformation.

[120] Michael Sailer and Lisa Homner. 2020. The gamification of learning: A meta-analysis. *Educational psychology review* 32, 1 (2020), 77–112.

[121] Joni Salminen, Chang Liu, Wenjing Pian, Jianxing Chi, Essi Häyhänen, and Bernard J Jansen. 2024. Deus ex machina and personas from large language models: investigating the composition of AI-generated persona descriptions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.

[122] Lauren Scott, Lynne Coventry, Marta E Cecchinato, and Mark Warner. 2023. "I figured her feeling a little bit bad was worth it to not spread that kind of hate": Exploring how UK families discuss and challenge misinformation. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–15.

[123] Joongi Shin, Michael A. Hedderich, Bartłomiej Jakub Rey, Andrés Lucero, and Antti Oulasvirta. 2024. Understanding Human-AI Workflows for Generating Personas. In *Proceedings of the Designing Interactive Systems Conference (DIS '24)* (Copenhagen, Denmark). Association for Computing Machinery, New York, NY, USA, 25 pages. https://doi.org/10.1145/3643834.3660729

[124] Jieun Shin, Lian Jian, Kevin Driscoll, and François Bar. 2018. The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior* 83 (2018), 278–287.

[125] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter* 19, 1 (2017), 22–36.

[126] Ismar Frango Silveira. 2016. Open educational games: Challenges and perspectives. In *2016 XI Latin American Conference on Learning Objects and Technology (LACLO)*. IEEE, IEEE, 1–9.

[127] Caroline Sinders. 2022. From Our Fellows: The Use of Mis- and Disinformation in Online Harassment Campaigns — cdt.org. https://cdt.org/insights/the-use-of-mis-and-disinformation-in-online-harassment-campaigns/. [Accessed 29-07-2024].

[128] Zijing Song, Yating Sun, and Ray Lc. 2022. DRIZZLE: A Comic for Covert Climate Action Influence. In *[ ] With Design: Reinventing Design Modes*, Gerhard Bruyns and Huaxin Wei (Eds.). Springer Nature, Singapore, 1613–1623. https://doi.org/10.1007/978-981-19-4472-7_105

[129] Zijing Song, Yating Sun, Vincent Ruijters, and RAY LC. 2021. Climate Influence: Implicit Game-Based Interactive Storytelling for Climate Action Purpose. In *Interactive Storytelling*, Alex Mitchell and Mirjam Vosmeer (Eds.). Springer International Publishing, Cham, 425–429. https://doi.org/10.1007/978-3-030-92300-6_42

[130] Neeraj Soni. 2024. 'Prebunking': An Effective Approach to Combat Misinformation. CyberPeace. https://www.cyberpeace.org/resources/blogs/prebunking-an-effective-approach-to-combat-misinformation

[131] Brian G Southwell, Jeff Niederdeppe, Joseph N Cappella, Anna Gaysynsky, Dannielle E Kelley, April Oh, Emily B Peterson, and Wen-Ying Sylvia Chou. 2019. Misinformation as a misunderstood challenge to public health. *American journal of preventive medicine* 57, 2 (2019), 282–285.

[132] Tobia Spampatti, Ulf JJ Hahnel, Evelina Trutnevyte, and Tobias Brosch. 2024. Psychological inoculation strategies to fight climate disinformation across 12 countries. *Nature Human Behaviour* 8, 2 (2024), 380–398.

[133] Victor Suarez-Lledo and Javier Alvarez-Galvez. 2021. Prevalence of health misinformation on social media: systematic review. *Journal of medical Internet research* 23, 1 (2021), e17187.

[134] Briony Swire-Thompson, David Lazer, et al. 2020. Public health and online misinformation: challenges and recommendations. *Annu Rev Public Health* 41, 1 (2020), 433–451.

[135] Edson C Tandoc Jr, Andrew ZH Yee, Jeremy Ong, James Chong Boi Lee, Duan Xu, Zheng Han, Chew Chee Han Matthew, Janelle Shaina Hui Yi Ng, Cui Min Lim, Lydia Rui Jun Cheng, et al. 2021. Developing a perceived social media literacy scale: Evidence from Singapore. *International Journal of Communication* 15 (2021), 22.

[136] Huiyun Tang, Gabriele Lenzini, Samuel Greiff, Björn Rohles, and Anastasia Sergeeva. 2024. "Who Knows? Maybe it Really Works": Analysing Users' Perceptions of Health Misinformation on Social Media. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA, 1499–1517.

[137] Haoheng Tang and Mrinalini Singha. 2024. A Mystery for You: A fact-checking game enhanced by large language models (LLMs) and a tangible interface. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM New York, NY, New York, NY, 1–5.

[138] Sigal Tifferet. 2021. Verifying online information: Development and validation of a self-report scale. *Technology in Society* 67 (2021), 101788.

[139] Sigmund Tobias, J Dexter Fletcher, and Alexander P Wind. 2014. Game-based learning. *Handbook of research on educational communications and technology* (2014), 485–503.

[140] Muhtar Çağkan Uludağlı and Kaya Oğuz. 2023. Non-player character decision-making in computer games. *Artificial Intelligence Review* 56, 12 (2023), 14159–14191.

[141] Jussi Valtonen, Ville-Juhani Ilmarinen, and Jan-Erik Lönnqvist. 2023. Political orientation predicts the use of conventional and complementary/alternative medicine: A survey study of 19 European countries. *Social Science & Medicine* 331 (2023), 116089.

[142] Gaurav Verma, Ankur Bhardwaj, Talayeh Aledavood, Munmun De Choudhury, and Srijan Kumar. 2022. Examining the impact of sharing COVID-19 misinformation online on mental health. *Scientific Reports* 12, 1 (2022), 8045.

[143] Emily Vraga, Melissa Tully, John E Kotcher, Anne-Bennett Smithson, and Melissa Broeckelman-Post. 2015. A Multi-Dimensional Approach to Measuring News Media Literacy. *Journal of media literacy education* 7, 3 (2015), 41–53.

[144] Yuxi Wang, Martin McKee, Aleksandra Torbica, and David Stuckler. 2019. Systematic literature review on the spread of health-related misinformation on social media. *Social science & medicine* 240 (2019), 112552.

[145] Zhilin Wang, Yu Ying Chiu, and Yu Cheung Chiu. 2023. Humanoid Agents: Platform for Simulating Human-like Generative Agents. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System*

Demonstrations, Yansong Feng and Els Lefever (Eds.). Association for Computational Linguistics, Singapore, 167–176. https://doi.org/10.18653/v1/2023.emnlp-demo.15

[146] Garrison Wells, Agnes Romhanyi, and Alaina Klaes. 2024. Sus: Modifying Among Us for Misinformation Discernment. In *Proceedings of the 19th International Conference on the Foundations of Digital Games (FDG 2024)* (Worcester, MA, USA). Association for Computing Machinery, New York, NY, USA, 4 pages. https://doi.org/10.1145/3649921.3656990

[147] Garrison Akira Wells. 2024. *DoomScroll: Modding Among Us to Combat Misinformation.* Ph. D. Dissertation. University of California, Irvine.

[148] Michael J Wood. 2018. Propagating and debunking conspiracy theories on Twitter during the 2015–2016 Zika virus outbreak. *Cyberpsychology, behavior, and social networking* 21, 8 (2018), 485–490.

[149] Liang Wu, Fred Morstatter, Kathleen M Carley, and Huan Liu. 2019. Misinformation in social media: definition, manipulation, and detection. *ACM SIGKDD explorations newsletter* 21, 2 (2019), 80–90.

[150] Chengxing Xie, Canyu Chen, Feiran Jia, Ziyu Ye, Kai Shu, Adel Bibi, Ziniu Hu, Philip Torr, Bernard Ghanem, and Guohao Li. 2024. Can Large Language Model Agents Simulate Human Trust Behaviors? *arXiv preprint arXiv:2402.04559* (2024).

[151] Yuzhuang Xu, Shuo Wang, Peng Li, Fuwen Luo, Xiaolong Wang, Weidong Liu, and Yang Liu. 2023. Exploring large language models for communication games: An empirical study on werewolf. *arXiv preprint arXiv:2309.04658* (2023).

[152] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, Toby Jia-Jun Li, and RAY LC. 2022. AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the IUI 2022 Workshops: APEx-UI, HAI-GEN, HEALTHI, HUMANIZE, TExSS, SOCIALIZE*. CEUR-WS Team, Aachen, Germany, 56–65. https://scholars.cityu.edu.hk/en/publications/publication(d901f5a2-0600-422f-b588-db5a59871961).html

[153] Soeun Yang, Jae Woo Lee, Hyoung-Jee Kim, Minji Kang, EunRyung Chong, and Eun-mee Kim. 2021. Can an online educational game contribute to developing information literate citizens? *Computers & Education* 161 (2021), 104057.

[154] Michael Yin, Emi Wang, Chuoxi Ng, and Robert Xiao. 2024. Lies, Deceit, and Hallucinations: Player Perception and Expectations Regarding Trust and Deception in Games. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)* (Honolulu, HI, USA). Association for Computing Machinery, New York, NY, USA, 15 pages. https://doi.org/10.1145/3613904.3642253

[155] Sami R Yousif, Rosie Aboody, and Frank C Keil. 2019. The illusion of consensus: A failure to distinguish between true and false consensus. *Psychological Science* 30, 8 (2019), 1195–1204.

[156] Yuhan Zeng, Yingxuan Shi, Xuehan Huang, Fiona Nah, and RAY LC. 2025. "Ronaldoâ€™s a poser!": How the Use of Generative AI Shapes Debates in Online Forums. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 18. https://doi.org/10.1145/3706598.3713829

[157] Qinshi Zhang, Ruoyu Wen, Latisha Besariani Hendra, Zijian Ding, and RAY LC. 2025. Can AI Prompt Humans? Multimodal Agents Prompt Playersâ€™ Game Actions and Show Consequences to Raise Sustainability Awareness. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 29. https://doi.org/10.1145/3706598.3713661

[158] Xichen Zhang and Ali A Ghorbani. 2020. An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management* 57, 2 (2020), 102025.

[159] Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. 2024. Memorybank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. AAAI Press, Palo Alto, CA, USA, 19724–19731.

[160] Suifang Zhou, Kexue Fu, Huamin Yi, and RAY LC. 2025. RetroChat: Designing for the Preservation of Past Chinese Online Social Experiences. In *Creativity and Cognition (C&C '25)*. Association for Computing Machinery, New York, NY, USA, 19. https://doi.org/10.1145/3698061.3726920

[161] Suifang Zhou, Latisha Hendra, and Ray Lc. 2024. Eternagram: Post-Climate Devastation Text Adventure. In *Proceedings of the 17th International Symposium on Visual Information Communication and Interaction (VINCI '24)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3678698.3687201

[162] Suifang Zhou, Latisha Besariani Hendra, Qinshi Zhang, Jussi Holopainen, and RAY LC. 2024. Eternagram: Probing Player Attitudes in Alternate Climate Scenarios Through a ChatGPT-Driven Text Adventure. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)* (Honolulu, HI, USA). Association for Computing Machinery, New York, NY, USA, 23 pages. https://doi.org/10.1145/3613904.3642850

[163] Suifang Zhou and Ray Lc. 2025. Eternagram: Inspiring Climate Action Through LLM-based Conversational Exploration of a Post-Devastation Climate Future. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3719160.3735658

Received 2025-02-19; accepted 2025-07-24