ImaginAItion: Promoting Generative AI Literacy Through Game-Based Learning

Qianou Ma
Human-Computer-Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
qianoum@cs.cmu.edu

Anika Jain*
Human-Computer-Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
anikajai@andrew.cmu.edu

Jini Kim*
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
jinik@andrew.cmu.edu

Megan Chai*
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
mvchai@andrew.cmu.edu

Geoff Kaufman
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
gfk@cs.cmu.edu

Abstract

The rapid adoption of Generative AI (GenAI) in our society and everyday life demonstrates the need for greater AI literacy of its potential biases and harms. Although there have been attempts to bring AI literacy to children, commonly via game-based learning, there is still a lack of instruction that aims to encourage a more nuanced understanding of the utility and harm of GenAI systems to a broader audience. To address the gap, we developed the educational game ImaginAItion, inspired by existing popular party game mechanisms such as Telestrations and Caution Signs. In particular, our game targets adults who do not have a deep understanding of GenAI. Leveraging persuasive strategies grounded in psychological theories, we seek to encourage deeper reflection on players' GenAI prompting behaviors and their understanding of its capabilities and limitations.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**; • **Applied computing** → **Computer-assisted instruction**; • **Computing methodologies** → Natural language generation.

Keywords

AI literacy, game-based learning, Generative AI

ACM Reference Format:
Qianou Ma, Anika Jain, Jini Kim, Megan Chai, and Geoff Kaufman. 2025. ImaginAItion: Promoting Generative AI Literacy Through Game-Based Learning. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3706599.3719844

*Equal contributions.

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3719844

1 Introduction

As Generative AI (GenAI) becomes increasingly integrated into diverse domains — from creative content generation to everyday decision support — there is a growing need to ensure that users not only learn how to operate GenAI but also develop a nuanced understanding of their capabilities and limitations [19, 22]. Specifically, the ability to critically evaluate GenAI outputs, recognize inherent biases and potential harms in GenAI, and form accurate mental models of how outputs are generated has emerged as a key component of AI literacy [9, 10, 27, 31].

While there is a growing body of HCI research aimed at helping adult learners understand complex AI algorithms [6, 18, 35, 37, 45] and a growing recognition of the importance of developing AI literacy for adult learners [11, 24], the field of GenAI literacy-related education for adult novices who do not have a deep understanding of GenAI remains in its early stages [2, 25, 28]. This highlights an urgent need for innovative educational methodologies to address these gaps and better support adult learners in navigating the complexities of GenAI systems [25, 28].

Game-based learning is popular in AI literacy interventions for children [2, 7], as games have the potential to provide scenarios that encourage experimentation, reflection, and engagement [26, 30, 43]. Beyond teaching procedural skills (e.g., crafting prompts), games may also foster critical awareness of biases and limitations inherent in GenAI, offering a more in-depth learning experience. However, limited work is focused on game-based learning approaches for adults to promote GenAI literacy, especially in recognizing biases and understanding the impacts of GenAI (details in Section 2).

In this work, we introduce ImaginAItion, a game-based intervention designed to help adult novices engage more deeply and critically with GenAI.[1] We position ImaginAItion as an exploratory

[1] In ImaginAItion, we instantiate "GenAI" with text-to-image GenAIs like DALL-E. More possibilities are discussed in Section 4.3.

environment where participants can develop GenAI literacy by interacting with generative models, iteratively refining their prompts, and reflecting on the nature of AI-generated content. Central to this inquiry are two research questions:

RQ.1 **Perceptions and Behaviors with GenAI**: How do adult novices understand bias in GenAI, perceive GenAI capacities, and prompt GenAI?

RQ.2 **GenAI Literacy Learning**: Can ImaginAItion teach GenAI literacy to adult novices, and specifically, help them develop more accurate GenAI bias awareness and mental model and successfully and efficiently produce and improve prompts?

We iteratively design the ImaginAItion game[2] (Section 3) with detailed learning objectives (Section 3.2) and plan to conduct an empirical study (Section 4). We aim to explore how playfully structured interactions with GenAI might support critical thinking, influence prompt engineering skills, and ultimately foster more informed and responsible relationships with GenAI among adult novices.

2 Related Works

2.1 AI and GenAI Literacy and Instruction

AI literacy refers to the ability to understand how AI operates, effectively utilize it, and critically evaluate the outputs of AI systems [27]. With the rise of GenAI tools such as ChatGPT and DALL-E, it is essential to cultivate GenAI literacy. Adults frequently encounter GenAI systems in both personal and professional contexts, making them particularly vulnerable to issues such as overreliance and misuse [11, 24]. However, research on GenAI literacy instructions, especially for adult novices, remains in its early stages [2, 25, 29].

Existing approaches to AI literacy education have predominantly focused on younger audiences, especially K-12 students [2, 7]. Although these initiatives are valuable for introducing foundational AI principles, they may fail to address the unique learning needs of adults who interact with AI in more complex and nuanced ways [25]. Methods such as workshops, tutoring systems, and games have shown promise in introducing foundational concepts of AI literacy [6, 7, 12]. However, current approaches often cover a limited range of topics, such as bias in training data or AI mechanisms visualizations [18, 35, 37, 45], or require substantial time and resources, making them less accessible to broader audiences. Additionally, GenAI literacy instruction would need to extend on traditional AI literacy instruction and address unique aspects of Gen AI, including understanding prompt engineering mechanisms, limitations, and inherent biases in GenAI [1, 3, 4, 17].

To address these gaps, we introduce the ImaginAItion game, a lightweight and engaging approach to GenAI literacy education that targets adult novices while addressing a broader range of learning objectives (see Section 3.2). We aim to design a time-efficient and scalable game, providing a more comprehensive GenAI literacy experience to users.

2.2 AI Literacy Game and Game-based Learning

An increasing body of research highlights the potential of games as effective tools to improve AI literacy [12, 13, 32, 47, 48]. For example, Zammit et al. [48] introduced TreasureIsland, which gamifies eBooks to improve AI literacy by effectively boosting students' motivation, self-efficacy, and understanding of AI concepts. Existing research shows that games have the potential to provide scenarios that encourage experimentation, reflection, and engagement, making them an effective environment for exploring the limitations and biases of AI systems [26, 30, 43]. Previous research has also emphasized that play serves as a valuable context for exploring and evolving users' mental models of AI systems [41, 42], and has suggested that games can help players critically understand the weaknesses of AI systems [40].

Furthermore, games utilize psychological theories and persuasive strategies to achieve their learning goals. For example, Kaufman et al. [21] adopted embedded design methods [20] like obfuscation and intermixing to make players more receptive to potentially threatening content like cross-gender role plays, and Tikka et al. [39] encourage more deliberate reflection and foster healthy eating behavioral change in players using dual-process theory [14].

We believe that beyond exposing users to procedural skills like crafting prompts, games can foster a critical understanding of the biases and limitations embedded in GenAI, offering learners a more comprehensive and immersive educational experience for crucial GenAI literacy skills. However, no existing games explicitly aim to enhance literacy in GenAI. Pictionary-based games can engage a wide range of people as popular party games [8]. These games can also demonstrate AI capabilities and limitations, such as Google's Quick, Draw!, which has a machine learning model that guesses human drawings [16]. The multiplayer varieties, such as Telestrations [33] and Caution Signs [34], show the potential to promote reflective practices, as they usually have a scoring and discussion session after each round. However, these popular games have yet to be explicitly adapted as tools to improve GenAI literacy.

Building on these insights, we restructured the mechanics of Pictionary-inspired party games to create ImaginAItion, in which players prompt GenAI to generate images and observe potentially biased outputs during gameplay (Section 3.1). We carefully designed ImaginAItion grounded in persuasive theories and our learning goals, aiming to both understand and improve adult novices' GenAI literacy skills during gameplay.

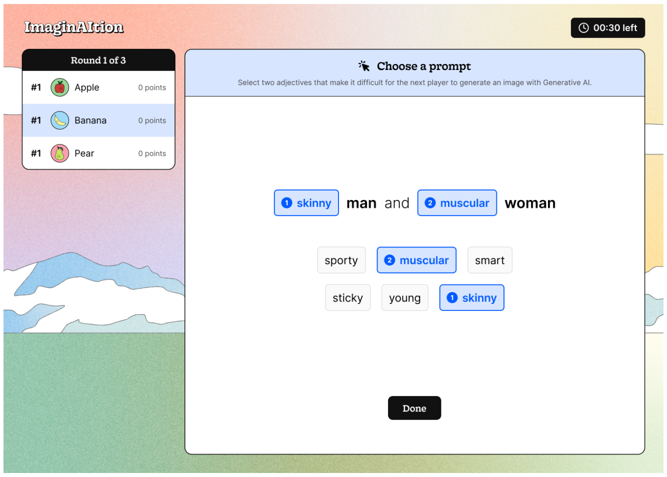
3 The Design of ImaginAItion

We conducted multiple iterations of prototyping for ImaginAItion to collect feedback and ensure that our game mechanics and interface (Section 3.1) are usable and engaging. We first defined our target audience and potential learning goals and developed multiple game concepts with internal playtesting. We designed low-fidelity prototypes combining multiple modalities, including paper, existing physical games (Telestrations [33], Caution Signs [34], and Pictionary [8]), and digital GenAI apps like Meta AI. We adopted embedded design [15] and dual-process theory [14] to best achieve practical learning goals and create an engaging game experience for players (Section 3.2). We also iterated on game mechanisms via external playtesting (Section 3.3).

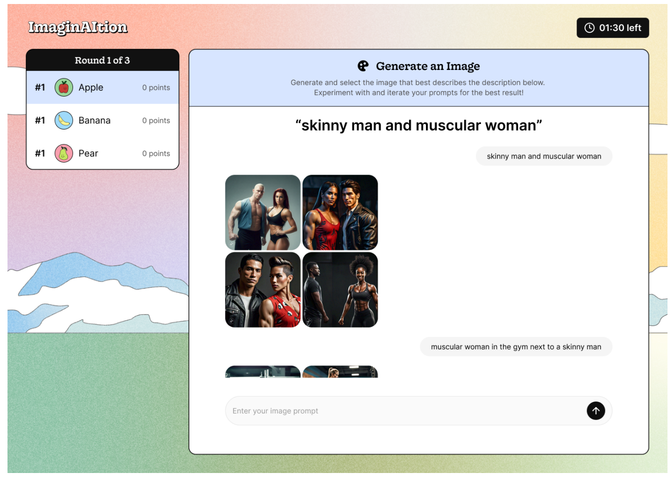
3.1 Game Overview

ImaginAItion [2] is a web-based multiplayer game in which three or more players alternate turns creating image prompts, drawing

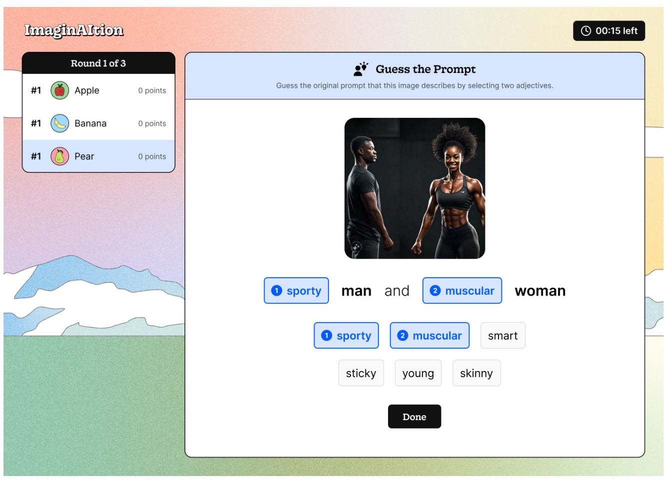
[2]Here is our interactive Figma prototype.



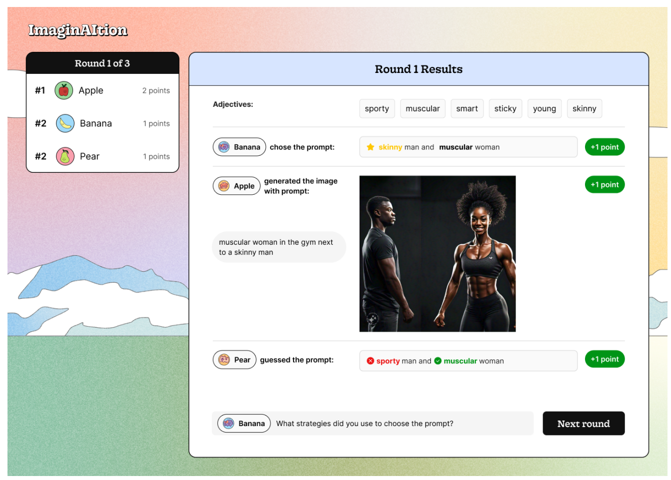
(a) Prompting turn. All players are presented with a set of six adjectives to select from. Players as prompters choose two adjectives to associate with the given nouns, and intentionally make it hard for the drawers to generate.



(b) Drawing turn. All players receive the prompt from the prompter and use GenAI to make images that describe the prompt. Players as drawers iterate using a chat interface and select an image to pass to the guessers by clicking on it.



(c) Guessing turn. All players as guessers can choose two adjectives out of the six presented that best describe the given image. This is the same adjective set that the prompter originally received.



(d) Scoring and Reflection. The prompter receives one point per adjective *not* identified by the guesser (e.g., "skinny"). The drawer and guesser receive one point per correct adjective (e.g., "muscular"). Players discuss strategies with a starter reflection question.

Figure 1: An example sequence in a round of ImaginAItion gameplay.

with GenAI tools, or guessing the image prompt to encourage greater GenAI literacy. As discussed in Section 2.2, we designed ImaginAItion as a game to enable users to effectively reflect on GenAI biases without explicit instruction and to make the subject matter more approachable. Players are assigned to one of three teams, and the team with the most points accumulated during all rounds wins the game. Each round consists of three sequences (one per team): prompting, drawing, and guessing, followed by the scoring and reflection for the round. One complete sequence is illustrated below with Figure 1:

(1) **Prompting Turn** (Figure 1a; 40 seconds): For the first turn, the team is tasked with selecting two adjectives from a set of six and pairing them with two predefined nouns for the round. Each team sees a different set of six adjectives, but all three teams share the two nouns in a round. We design rounds that include words that can reveal biases in GenAI [44, 49] with nouns like "man" and "woman" and adjectives such as "skinny" and "muscular".[3] The prompter's goal is to create a challenging image prompt for the drawer to generate and the guesser to guess correctly.
(2) **Drawing Turn** (Figure 1b; 60 seconds, excluding image generation time): For the second turn, the player receives the selected

[3] A prompt such as "skinny man and muscular woman" has been hard to generate with current GenAI, as it reveals rooted bias in body type [44, 49]. We also mix bias-inducing prompts with off-topic playful prompts such as "lazy horse and innovative billionaire" (elaborated in Section 3.2).

prompt from the prompter and is tasked with generating an image that best depicts the adjective and noun pairings. Using GenAI, drawers are encouraged to iteratively refine the image prompt to generate an image that clearly aligns with the adjectives and select one image to pass to the next player, the guesser. The drawer's goal is to generate an accurate image so that the guesser can guess the adjective and noun pairing prompt correctly.

(3) **Guessing Turn** (Figure 1c; 20 seconds): For the third turn, the player evaluates the image passed to them from the drawer and is tasked to guess the original image prompt from a set of six provided adjectives (same as what the prompter saw). The guesser's goal is to select the two adjectives that best describe the given image.

(4) **Scoring and Reflection** (Figure 1d; untimed): Scoring occurs when a round is completed, with two possible points in each prompt → draw → guess turn sequence. The prompter is awarded points for adjectives that were incorrectly guessed, as their objective was to make it difficult for the drawer to generate. The drawer and the guesser are awarded points for adjectives that the guesser correctly identified. All players review the three sequences along with the scores and discuss strategies and observations, using provided reflection questions as a starting point.

3.2 Learning Objectives and Key Design Rationale

Overall, our goal is to promote a deeper understanding of GenAI and to center the mental models of GenAI users around both the utility and harm of GenAI systems (aligned with existing AI literacy framework [27]). We define our three specific learning objectives as follows, and examples of learning outcome measurements are presented in Table 1 (full details in Appendix A.2).

After interacting with ImaginAItion, users should be able to:

LO1 **Bias Awareness**: identify and describe specific examples of bias in GenAI.

LO2 **Mental Model**: produce and assess specific examples of what is easy or hard for GenAI.

LO3 **Prompt Engineering**: successfully and efficiently produce and improve prompts for GenAI.

We highlight three most essential design rationales for ImaginAItion derived from our iterative design:

Highlight GenAI literacy goals with purposeful practice and feedback loop. ImaginAItion is designed to provide players with meaningful opportunities to practice key GenAI literacy skills while receiving actionable feedback. Prompters are motivated by the scoring system to select adjectives that would be difficult for the drawers, which encourages them to reflect on the potential biases (LO1) and limitations (LO2) in GenAI. Using GenAI in image generation ensures that players actively practice prompting (LO3) and experience the capabilities and limitations of GenAI and prompts (LO2) first-hand. Finally, a scoring session (Figure 1d) provides players with feedback on their performance, strengthening their understanding of how GenAI interprets and executes prompts. Discussion questions encourage learning from each other and deeper reflections on GenAI biases (LO1; e.g., "what bias might the generation reflect?"), capacities (LO2; e.g., "what strategies did you use to choose a difficult prompt?"), and prompting strategies (LO3; e.g., "what strategies did you use to prompt image generation?"), which foster learning throughout different rounds of the game.

Improve learning through embedding persuasive theories. We leverage persuasive theories such as obfuscation and intermixing in embedded design [15] and fast and slow thinking in dual-process theory [14] to maximize learning impact. Embedded design principles can make discussions on Gen AI bias more approachable, especially to those without subject matter knowledge [15]. Potentially triggering bias-related concepts are subtly integrated into gameplay through obfuscation, allowing players to reflect on biases without overt instruction that might reduce enjoyment and learning in games [23]. We designed adjective and noun combinations to highlight potential biases in GenAI outputs, and bias-revealing rounds are intermixed with purely playful prompts to maintain engagement while promoting critical thinking. Additionally, time constraints in each turn of the game are carefully designed so that the guessers need to rely more on their instincts when making automatic judgments matching prompts with images, while the prompters and the drawers are given more time for deliberate, logical reasoning and iterations [14]. The scoring and reflection questions also allow players to refine their mental models of GenAI, helping them develop the critical thinking skills necessary to engage with GenAI responsibly.

Reduce cognitive load by simplifying game mechanisms. To ensure ImaginAItion is lightweight and appeals to a wide audience of adult novices, we intentionally streamlined its mechanics to minimize cognitive load while retaining depth and engagement. For example, players create prompts and make guesses using predefined options in a consistent multi-select interface, eliminating the complexity of generating and evaluating open-ended input. An intuitive scoring interface further reduces confusion by clearly displaying how points are earned across different stages of the game. This design ensures that players can easily track their progress and understand the relationship between their decisions, the outputs generated by GenAI, and the resulting scores. We aim to create a supportive and engaging environment where players can explore challenging topics such as bias without feeling overwhelmed.

3.3 Playtest Insights

3.3.1 Scaffolded Prompting. Early iterations of ImaginAItion explored free-form prompts, which we found difficult to ideate, guess, and evaluate, and also risk digressing away from GenAI literacy goals to a vocabulary test. Therefore, we introduced a structured prompt format pairing adjectives with two predefined nouns, as illustrated in Figure 1a. We took inspiration from the Caution Signs game but increased the number of nouns and adjectives to better expose biases and limitations in GenAI (LO1, LO2), as we found that GenAI had challenges generating images of noun pairings that reflect bias such as "skinny man and muscular woman". Through experimentation, we found that varying nouns or verbs between nouns were too obvious or added unnecessary complexity, while fixed nouns with varied adjectives kept images relevant for discussion and aligned with our learning goals. To balance learning and fun, we intermixed bias-inducing prompts, such as comparing "man" and "woman" to highlight social bias ingrained in GenAI, with random playful prompts like "horse" and "billionaire" to keep the game engaging.

3.3.2 Role of GenAI. We tested multiple roles and interactions for GenAI in ImaginAItion, including GenAI always drawing or guessing, mixing human player and GenAI turns randomly, GenAI assisting human player decisions, and GenAI co-participating alongside human players. Overall, we found that player engagement dropped when GenAI took over the guessing roles, and prompt complexity became difficult to balance when both GenAI and human players drew or guessed. For example, when we tested having some players act as a GenAI agent that guesses and draws, and all other players draw and guess by hand, we found it difficult to create prompts that were complex enough to challenge both GenAI and humans. Through iterations, we found that letting humans iterate and pick GenAI results was the most effective in allowing players to practice refining prompts (LO3), and having only humans guess prompts based on images was the most straightforward way to evaluate prompts and reveal potential instances of GenAI bias.

3.3.3 Player Involvement. We explored the dynamics around each player's role in the game to best encourage engaging play. In the initial prototype, we tested having one player as the guesser, similar to Caution Signs. However, as we used GenAI to generate images and had a more complex prompt structure, the guesser was uninvolved for an extended time. Therefore, we adopted a turn-based structure similar to Telestrations, where players are always engaged in choosing prompts, drawing based on a given prompt, or guessing based on the drawings. The design of three teams and three turns per round keeps the game engaging and provides ample opportunities to practice GenAI literacy goals. More setups can be explored in future work (Section 4.3).

3.3.4 GenAI Literacy Learning. In our playtesting, we found preliminary evidence that our design fosters active GenAI literacy learning during gameplay. We found that our prompt structure allowed players to reflect on biases in GenAI (LO1) while maintaining the flow of the game, as the participants discussed the embedded bias in gender for the round exemplified in Figure 1 and found the game very fun. Interestingly, participants commented that they became aware of their own biases or purposefully used stereotypes to communicate efficiently — for example, instructing GenAI to add a ribbon to the cow to communicate the prompt "beautiful cow". Participants also improved their understanding of GenAI limitations (LO2), as those who struggled to create challenging prompts initially could instantiate scenarios GenAI handles poorly after gameplay. Additionally, participants demonstrated improved prompting strategies (LO3). Before the gameplay, participants described that they simply "regenerate the image again and again" and "retype prompt and try again" when GenAI fails; after gameplay, they derived more effective approaches such as "write more about the structure of the image" or "rethink and try to simplify or add more context".

4 Next Step: Game Evaluation Study

We will implement ImaginAItion as a web application with OpenAI's DALL·E 3 API for the GenAI model, and we plan to evaluate the efficacy of our game and answer our research questions in Section 1 with a within-subject study design.

4.1 Participants and Study Procedure

We plan to recruit 20 adult novices who use GenAI tools like ChatGPT but may lack a deep understanding of GenAI technologies. We will use a screening survey (Appendix A) to exclude participants who have high self-rated familiarity with GenAI (> 4/7 in 7-level Likert Scale questions).

The study will be conducted in 50-minute Zoom sessions, each involving three participants to meet the game requirements. Participants will be compensated with a $15 Gift Card for participating in the study. Please refer to Appendix A for our complete study materials. Each session will include the following phases:

- **Pre-survey** (5 minutes): Participants will provide demographic information, exemplify GenAI biases and prompting strategies, and rate their experiences with GenAI tools (e.g., confidence using tools like DALL·E) with 7-level Likert Scale questions. Table 1 provides a mapping between survey items and ImaginAItion's learning objectives, and complete details are in Appendix A.2.
- **Tutorial and Gameplay** (30 minutes): Participants will receive a tutorial on how to play ImaginAItion, with fixed bias-inducing prompts in the prompting turn for priming [36], and without the time constraints to enable participants to familiarize themselves with the game mechanics. Participants will then play five full rounds of ImaginAItion web game (5 minutes each, as described in Section 3.1), two of which include bias-inducing adjectives, and the other three are random.
- **Post-survey and Discussion** (15 minutes): After the gameplay, participants will complete a post-survey with a similar set of open-ended and Likert Scale questions question to the pre-survey to measure changes in their understanding of GenAI biases, mental models, and prompting strategies. This will be followed by a semi-structured discussion facilitated by the researchers, to collect qualitative insights on participants' experiences and reflections.

4.2 Method and Data Collection

To understand participants' experiences, prompting behaviors, and learning outcomes, we will collect data including participants' survey answers, time on task, and interaction log data from all phases of the study. As presented in Table 1, some pre-post survey items are specifically designed to evaluate ImaginAItion's learning objectives using backward design [46], a popular instructional design method to align educational goal, assessment, and instruction.

Data Analysis. We will use thematic analysis [5] and grounded theory coding [38] on the survey responses, discussion scripts, and interaction logs to derive qualitative insights regarding participants' gameplay patterns, mental models, and understanding of bias toward GenAI. We will use AI-assisted expert judgment to evaluate the open-ended answers in the pre-post survey.

Table 1: Alignment among the learning objectives, assessment (pre-/post-survey questions and in-game metrics), and game mechanisms in ImaginAItion. Refer to Appendix A for more survey questions.

LO	Assessment Example	Game Mechanism
LO1 (Bias Awareness)	Bias-inducing adjective selection strategies during gameplay; survey and interview questions on bias (e.g., *"What bias did you observe from the gameplay?"*)	Prompters may manipulate GenAI bias when selecting prompts. Drawers may experience GenAI bias when assigned a bias-inducing prompt and observe difficulties in GenAI generation.
LO2 (Mental Model)	Scores as prompters during gameplay; survey and interview questions on GenAI capacities (e.g., *"Write 3 prompts that you think would be challenging for text-to-image generative AI to draw"*) and prompt selection strategies (e.g., *"What selection strategies did you use to make the prompt hard?"*)	Prompters need to reflect on what is difficult for GenAI to generate, incentivized by the scoring mechanism. Drawers need to iteratively prompt and observe model behavior to refine their understanding of what GenAI generates well versus poorly.
LO3 (Prompt Engineering)	Prompt iterations, time, scores as drawers during gameplay; survey and interview questions on prompting strategies (e.g., *"Here's a prompt and its output for a text-to-image GenAI: [prompt that failed] [image]. Edit the prompt to make GenAI generate a better image."*)	Drawers need to iteratively prompt and observe model behavior to create images that effectively convey the selected adjectives by the prompter.

4.3 Future Work

To further reinforce the learning objectives, we may also adopt more complex game mechanisms, such as enabling the prompter to use their points to customize prompts or add distractor options for the guesser. Splitting more than three players into three teams encourages collaboration and communication, but it is also possible to explore individual gameplay similar to Telestrations, which keep repeating the drawing and guessing turns. Additionally, beyond the self-practice mechanisms with non-deterministic GenAI feedback, it would be interesting to introduce more explicit prompt engineering instructions in the gameplay, potentially borrowing insights from existing prompt engineering training literature [29]. Also note that in our work, we instantiate GenAI with the specific text-to-image GenAIs we tested when we developed our ImaginAItion game: Meta AI and DALL·E 3. The lessons learned may be generalizable to other GenAI models and LLMs, and potentially to other demographics like kids or families (as the recommended age for the game Caution Signs and Telestrations is 8+ and 12+ years, respectively), which we left for future work.

5 Conclusion

ImaginAItion is a party game that aims to promote GenAI literacy in adult novices who do not deeply understand GenAI technologies. By integrating principles from psychological theories, the game is designed to provide players with hands-on experiences to identify biases, refine their understanding of GenAI systems, and practice prompting. Looking ahead, the evaluation study will validate the game's effectiveness in achieving its learning objectives, offering insights into how structured gameplay can foster a deeper awareness of GenAI's capabilities and limitations. By expanding this research to include broader demographics and continuing to evolve the mechanics of the game, ImaginAItion could contribute to empowering users to navigate the complexities of GenAI responsibly.

Acknowledgments

Thanks to all the participants in this work. Thanks to Sherry Tongshuang Wu, Hong Shen, Erik Harpstead, Isadora Krsek, and other classmates from the CMU F24 Persuasive Design class for all the insights and feedback. Funding to attend this conference was provided by the CMU GSA/Provost Conference Funding.

References

[1] Haesol Bae and Aras Bozkurt. 2024. The Untold Story of Training Students with Generative AI: Are We Preparing Students for True Learning or Just Personalization? *Online Learning* 28, 3 (2024).

[2] Maalvika Bhat and Duri Long. 2024. Designing Interactive Explainable AI Tools for Algorithmic Literacy and Transparency. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 939–957. https://doi.org/10.1145/3643834.3660722

[3] Aras Bozkurt. 2024. Why generative AI literacy, why now and why it matters in the educational landscape?: Kings, queens and GenAI dragons. , 283–290 pages.

[4] Aras Bozkurt and Haesol Bae. [n. d.]. May the Force Be with You JedAI: Balancing the Light and Dark Sides of Generative AI in the Educational Landscape. ([n. d.]).

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[6] Michelle Carney, Barron Webster, Irene Alvarado, Kyle Phillips, Noura Howell, Jordan Griffith, Jonas Jongejan, Amit Pitaru, and Alexander Chen. 2020. Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3382839

[7] Lorena Casal-Otero, Alejandro Catala, Carmen Fernández-Morante, Maria Taboada, Beatriz Cebreiro, and Senén Barro. 2023. AI literacy in K-12: a systematic literature review. *International Journal of STEM Education* 10, 1 (April 2023), 1–17. https://doi.org/10.1186/s40594-023-00418-7

[8] Wikipedia contributors. [n. d.]. Pictionary. https://en.wikipedia.org/wiki/Pictionary. Accessed: December 8, 2024.

[9] Andrew Cox. 2024. Algorithmic literacy, AI literacy and responsible generative AI literacy. *Journal of Web Librarianship* (2024), 1–18.

[10] Leyla Dogruel, Philipp Masur, and Sven Joeckel. 2022. Development and validation of an algorithm literacy scale for internet users. *Communication Methods and Measures* 16, 2 (2022), 115–133.

[11] Stefania Druga, Nancy Otero, and Amy J. Ko. 2022. The Landscape of Teaching Resources for AI Education. In *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1* (Dublin, Ireland) *(ITiCSE '22)*. Association for Computing Machinery, New York, NY, USA, 96–102. https://doi.org/10.1145/3502718.3524782

[12] Xiaoxue Du, Zhichun Liu, Xi Wang, and Qiping Tang. 2024. Fostering AI Literacy through Interactive Game-Based Learning: A Case Study on Enhancing Algorithmic Thinking Skills. In *Society for Information Technology & Teacher Education International Conference*. Association for the Advancement of Computing in Education (AACE), 333–338.

[13] Xiaoxue Du and Xi Wang. 2023. Play by Design: Developing Artificial Intelligence Literacy through Game-based Learning. *Journal of Computer Science Research* 5, 4 (2023), 1–12.

[14] Jonathan St BT Evans and Keith E Stanovich. 2013. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science* 8, 3 (2013), 223–241.

[15] Mary Flanagan Geoff Kaufman. 2015. A psychologically "embedded" approach to designing games for prosocial causes. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* (2015).

[16] Google. [n. d.]. Quick, Draw! https://quickdraw.withgoogle.com/. Accessed: December 8, 2024.

[17] Anuj Gupta, Yasser Atef, Anna Mills, and Maha Bali. 2024. Assistant, parrot, or colonizing loudspeaker? ChatGPT metaphors for developing critical AI Literacies. *Open Praxis* 16, 1 (2024), 37–53.

[18] Minsuk Kahng, Nikhil Thorat, Duen Horng Polo Chau, Fernanda B. Viegas, and Martin Wattenberg. 2019. GAN Lab: Understanding Complex Deep Generative Models using Interactive Visual Experimentation. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan. 2019), 310–320. https://doi.org/10.1109/tvcg.2018.2864500

[19] Maria Kasinidou. 2024. Development of Personalised Educational Tools for AI Literacy Using Participatory Design. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization* (Cagliari, Italy) *(UMAP Adjunct '24)*. Association for Computing Machinery, New York, NY, USA, 30–34. https://doi.org/10.1145/3631700.3664917

[20] Geoffrey Kaufman and Mary Flanagan. 2015. A Psychologically "Embedded" Approach to Designing Games for Prosocial Causes. *Cyberpsychology, Behavior, and Social Networking* 18, 4 (2015), 241–248. https://doi.org/10.1089/cyber.2014.0513

[21] Geoff Kaufman, Mary Flanagan, and Gili Freedman. 2019. Not just for girls: Encouraging cross-gender role play and reducing gender stereotypes with a strategy game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, New York, NY, USA. https://doi.org/10.1145/3311350.3347177

[22] Patrick Gage Kelley, Yongwei Yang, Courtney Heldreth, Christopher Moessner, Aaron Sedley, Andreas Kramm, David T. Newman, and Allison Woodruff. 2021. Exciting, Useful, Worrying, Futuristic: Public Perception of Artificial Intelligence in 8 Countries. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)*. ACM. https://doi.org/10.1145/3461702.3462605

[23] Eric Klopfer, Scot Osterweil, and Katie Salen. 2009. Moving learning games forward. (2009).

[24] Valentin Kuleto, Milena Ilić, Mihail Dumangiu, Marko Ranković, Oliva MD Martins, Dan Păun, and Larisa Mihoreanu. 2021. Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. *Sustainability* 13, 18 (2021), 10424.

[25] Matthias Carl Laupichler, Alexandra Aster, Jana Schirch, and Tobias Raupach. 2022. Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence* 3, 100101 (Jan. 2022), 100101. https://doi.org/10.1016/j.caeai.2022.100101

[26] Andreas Lieberoth. 2015. Shallow gamification: Testing psychological effects of framing an activity as a game. *Games and Culture* 10, 3 (2015), 229–248.

[27] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–16.

[28] Duri Long, Jessica Roberts, Brian Magerko, Kenneth Holstein, Daniella DiPaola, and Fred Martin. 2023. Ai literacy: Finding common threads between education, design, policy, and explainability. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–6.

[29] Qianou Ma, Weirui Peng, Chenyang Yang, Hua Shen, Kenneth Koedinger, and Tongshuang Wu. 2024. What Should We Engineer in Prompts? Training Humans in Requirement-Driven LLM Use. *arXiv preprint arXiv:2409.08775* (2024). https://arxiv.org/abs/2409.08775

[30] Katelyn Morrison, Mayank Jain, Jessica Hammer, and Adam Perer. 2023. Eye into AI: Evaluating the Interpretability of Explainable AI Techniques through a Game with a Purpose. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 273 (Oct. 2023), 22 pages. https://doi.org/10.1145/3610064

[31] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. 2021. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence* 2, 100041 (Jan. 2021), 100041. https://doi.org/10.1016/j.caeai.2021.100041

[32] Davy Tsz Kit Ng, Chen Xinyu, Jac Ka Lok Leung, and Samuel Kai Wah Chu. 2024. Fostering students' AI literacy development through educational games: AI knowledge, affective and cognitive engagement. *Journal of Computer Assisted Learning* (2024).

[33] The Op. [n. d.]. Telestrations. https://theop.games/collections/telestrations. Accessed: December 8, 2024.

[34] Danielle Reynolds and Scott Brady. [n. d.]. CautionSigns. https://cautionsigns.app/. Accessed: December 8, 2024.

[35] Benjamin Sanchez-Lengeling, Emily Reif, Adam Pearce, and Alexander B Wiltschko. 2021. A gentle introduction to graph neural networks. *Distill* 6, 9 (2021), e33.

[36] Joseph Seering, Tianying Fang, Lila Damasco, Ming C. Chen, Liting Sun, and Geoffrey Kaufman. 2019. Designing User Interface Elements to Improve the Quality and Civility of Discourse in Online Commenting Behaviors. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300836

[37] Daniel Smilkov, Shan Carter, D. Sculley, Fernanda B. Viégas, and Martin Wattenberg. 2017. Direct-Manipulation Visualization of Deep Networks. arXiv:1708.03788 [cs.LG] https://arxiv.org/abs/1708.03788

[38] Robert Thornberg, Kathy Charmaz, et al. 2014. Grounded theory and theoretical coding. *The SAGE handbook of qualitative data analysis* 5, 2014 (2014), 153–69.

[39] Piiastiina Tikka, Miia Laitinen, Iikka Manninen, and Harri Oinas-Kukkonen. 2018. Reflection through gaming: Reinforcing health message response through gamified rehearsal. In *Persuasive Technology*. Springer International Publishing, Cham, 200–212. https://doi.org/10.1007/978-3-319-78978-1_17

[40] Jennifer Villareale, Ana V Acosta-Ruiz, Samuel Adam Arcaro, Thomas Fox, Evan Freed, Robert C Gray, Mathias Löwe, Panote Nuchprayoon, Aleksanteri Sladek, Rush Weigelt, Yifu Li, Sebastian Risi, and Jichen Zhu. 2020. INNk: A multi-player game to deceive a neural network. In *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*. ACM, New York, NY, USA. https://doi.org/10.1145/3383668.3419858

[41] Jennifer Villareale, Gabriele Cimolino, and Daniel Gomme. 2023. Playing with dezgo: Adapting human-AI interaction to the context of play. In *Proceedings of the 18th International Conference on the Foundations of Digital Games*. ACM, New York, NY, USA. https://doi.org/10.1145/3582437.3587198

[42] Jennifer Villareale, Casper Harteveld, and Jichen Zhu. 2022. "I want to see how smart this AI really is": Player mental model development of an adversarial AI player. *Proc. ACM Hum. Comput. Interact.* 6, CHI PLAY (Oct. 2022), 1–26. https://doi.org/10.1145/3549482

[43] Luis Von Ahn and Laura Dabbish. 2008. Designing games with a purpose. *Commun. ACM* 51, 8 (2008), 58–67.

[44] Adriana Fernández de Caleya Vázquez and Eduardo C Garrido-Merchán. 2024. A taxonomy of the biases of the images created by generative artificial intelligence. *arXiv [cs.CY]* (May 2024). arXiv:2407.01556 [cs.CY] http://arxiv.org/abs/2407.01556

[45] Zijie J Wang, Robert Turko, Omar Shaikh, Haekyu Park, Nilaksh Das, Fred Hohman, Minsuk Kahng, and Duen Horng Polo Chau. 2020. CNN explainer: learning convolutional neural networks with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 1396–1406.

[46] Grant P Wiggins and Jay McTighe. 2005. *Understanding by Design*. ASCD. https://play.google.com/store/books/details?id=N2EfKlyUN4QC

[47] Michelle Yavorskiy and Lydia Kim. 2024. MeadowMinds: An AI Literacy Game. (2024).

[48] Marvin Zammit, Iro Voulgari, Antonios Liapis, and Georgios N Yannakakis. 2021. The road to AI literacy education: from pedagogical needs to tangible game design. Academic Conferences International.

[49] Mi Zhou, Vibhanshu Abhishek, Timothy Derdenger, Jaymo Kim, and Kannan Srinivasan. 2024. Bias in Generative AI. *arXiv [econ.GN]* (March 2024). arXiv:2403.02726 [econ.GN] http://arxiv.org/abs/2403.02726

Figure 2: How to play interface design. Players can view the instructions on how to play the game with an overview of the prompting, drawing, and guessing turns along with the scoring mechanisms.

A User Study Materials

A.1 Screening Survey Questions

(1) On a scale of 1-7 (Likert), how familiar are you with text-to-image generative AI tools?
(2) Have you ever learned prompting methods to enhance the performance of text-to-image generative AI? If so, briefly describe what you've learned.

A.2 Pre-Post Survey and Discussion Questions

(1) On a scale of 1-7 (Likert), how comfortable are you with text-to-image generative AI tools?
LO3 Prompt Engineering:
(2) On a scale of 1-7 (Likert), how easy do you find it is to create prompts that guide text-to-image generative AI to generate your desired images?
(3) What strategies do you use to make text-to-image generative AI generate images more aligned with your goals? Provide examples.
(4) Here's a prompt and its output for a text-to-image GenAI: [prompt that failed] [image]. Edit the prompt to make GenAI generate a better image.
LO2 Mental Model:
(5) Write 3 prompts that you think would be easy for text-to-image generative AI to draw.
(6) Write 3 prompts that you think would be challenging for text-to-image generative AI to draw.
LO1 Bias Awareness:
(7) What biases do you think exist in text-to-image generative AI? List as many as you can.
(8) Select All Apply: What prompt(s) can produce an image like this: [image], [list of prompts that reflect GenAI bias] (and why)?

Post-survey only:

(1) **LO2 Mental Model**: What selection strategies did you use to make the prompt hard?
(2) **LO1 Bias Awareness**: What bias did you observe from the gameplay?
(3) How is your gameplay experience?
(4) What have you learned from this game, if any?
(5) If there's one thing you can change about this game, what would you change and why?
(6) Other comments, feedback?

Semi-structured group discussion: Start by asking the post-survey-only questions above.

B Interface Artifacts

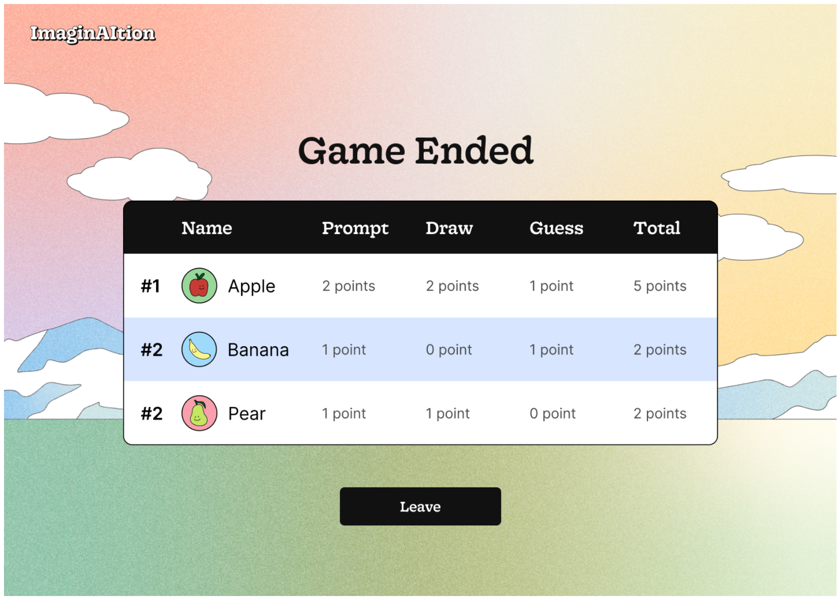


Figure 3: Final scoring interface design. Players can view the scores for each round and the final game scores.