Enhancing AI Explainability for Non-technical Users with LLM-Driven Narrative Gamification

Yuzhe You
School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
y28you@uwaterloo.ca

Helen Weixu Chen
School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
w352chen@uwaterloo.ca

Jian Zhao
School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
jianzhao@uwaterloo.ca

Abstract

Artificial intelligence (AI) is tightly integrated into modern technology, yet existing exploratory XAI visualizations are primarily designed for users with technical expertise. This leaves everyday users, who now also rely on AI systems for work and tasks, with limited resources to explore or understand AI. In this work, we explored the use of LLM-driven narrative gamification to enhance the learning and engagement of exploratory XAI visualizations. Specifically, we designed a design probe that enables non-experts to collect insights from an embedding projection by conversing directly with visualization elements similar to game NPCs. We conducted a preliminary comparative study to assess the effectiveness and usability of our design probe. Our study shows that while the tool enhances non-technical users' AI knowledge and is perceived as beneficial, the impact of gamification alone on understanding remains inconclusive. Participant opinions on engagement are mixed: some find it enriching, while others see it as disruptive.

CCS Concepts

• **Human-centered computing** → **Visualization**; **Human computer interaction (HCI)**.

Keywords

XAI Visualization, Explainable AI, Gamification

ACM Reference Format:

Yuzhe You, Helen Weixu Chen, and Jian Zhao. 2025. Enhancing AI Explainability for Non-technical Users with LLM-Driven Narrative Gamification. 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, 7 pages. https://doi.org/10.1145/3706599.3719795

1 Introduction

As artificial intelligence (AI) becomes deeply integrated into modern technology, it is transforming the lives of not just technical practitioners but also everyday people who increasingly interact with AI systems in their daily tasks [9, 18, 19, 36]. This growing interaction has created a need for non-technical users to also explore and understand AI systems, whether they are knowledge workers who rely on AI tools for work [7, 34], or AI enthusiasts simply seeking to stay informed about advancing technology [20]. Unlike traditional AI courses, which often demand significant time and cover broad topics, these users typically require hands-on, task-specific AI learning experiences tailored to their specific tasks and interests. Exploratory explainable AI (XAI) visualizations have been shown to effectively provide hands-on learning of the decision-making and performance of AI systems for experienced practitioners (e.g., [5, 15, 17, 37]). However, they have been criticized for presenting AI material in an overly technical way, making it challenging for everyday AI users to connect with and understand the content [6]. One promising yet underexplored approach to addressing this gap is gamification, the application of game-design elements in non-game contexts [10]. Gamification has been shown to effectively enhance learning and engagement by helping non-experts make meaningful connections with learning material [4, 12, 22, 24]. Despite its potential, gamification has not yet been explored in the context of supporting non-experts' understanding when exploring XAI visualizations. Moreover, recent advancements in Large Language Models (LLMs) present a new opportunity to take gamification a step further [1, 2, 33]. With their advanced reasoning and generation capabilities, LLMs can be leveraged in gamification strategies to create dynamic visualization exploration experience that adapts to each user's unique background and technical expertise. This work aims to explore how gamification, supported by LLMs, can transform exploratory XAI visualizations by making them more meaningful and engaging for everyday AI users.

While some existing works focus on gamifying tasks within XAI domain, their goal deviates from using gamification as a means to enhance users' understanding of AI when exploring XAI tools. For example, Fulton et al. [8] developed a game with a purpose (GWAP) where players guessed the source input image of a convolutional neural network (CNN) based on its feature visualizations, but this was designed to assess human interpretations rather than aid user comprehension. Sevastjanova et al. [31] incorporated gamification into a workspace for labeling question types as training data for supervised machine learning, which was also not designed to enhance AI understanding and primarily targeted experts rather than non-technical users. While these studies contribute to gamification in XAI, they do not explore its potential in gamifying XAI visualizations to improve users' AI understanding and engagement, a direction that is worth exploring especially with recent advancements in LLMs.

To explore how gamification could be used as a direction to augment XAI visualizations and improve learning and engagement for non-expert AI users, we investigate integrating gamification strategies powered by LLMs into existing XAI visualization techniques. In this paper, we present our initial exploration. Specifically, we

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3719795

examine the strategy of creating LLM-powered narrative gamification, where users can "speak" directly to NPC-like visualization elements in natural language to collect insights. This approach is motivated by findings in existing literature, which highlight that a significant barrier to AI learning for non-experts is the lack of relatedness [6, 39]. Non-experts often lack peers or mentors who can provide relatable support, leaving them without the social connections that foster learning. Our primary objective is to explore whether LLM-driven narrative gamification could promote an intuitive understanding of complex XAI visualizations and actively engage non-technical users in the exploration process. For this work, we created a design probe using an interactive t-SNE projection of image-based classifiers' embeddings, a widely used XAI visualization technique that helps identify clusters and anomalies to better understand model behavior and perceptions [15, 17, 35, 40]. To evaluate the usability and learning effects of the gamification, we conducted a preliminary between-subjects study comparing versions of the visualization with and without the LLM-based narrative gamification. We also plan to perform a qualitative study to gain deeper insights into how this gamification approach could serve as a promising direction to address the AI literacy gap for non-technical users exploring AI models through XAI visualizations. Our preliminary study results show that though our prototype effectively enhances non-technical users' AI/XAI knowledge, and that users believe they learn more through the gamification feature, it remains inconclusive whether the gamification itself leads to further improvements in understanding. Additionally, opinions among participants regarding the gamification's engagement are mixed: some believe the gamification enhances their exploration of the visualizations, while others feel that constantly having to converse with visualization elements disrupts their workflow. In summary, we make the following contributions:

- We present a preliminary **design study** that explores the direction of using LLM-driven narrative gamification to enhance the learning effects and engagement of exploratory XAI visualizations for non-technical AI users.
- We implemented a **design probe** that integrates LLM-powered narrative gamification into an interactive embedding projection, an existing popular XAI visualization. This design probe allows users to explore and understand an embedding projection by interacting with data points or cluster like they are game NPCs in a conversational manner.
- We perform a **user study** with 10 non-technical users to quantitatively and qualitatively assess the effects and usability of our prototype. We also use our design probe to extract initial empirical findings on how LLM-supported narrative gamification can help non-technical users better understand and enjoy XAI visualizations, identify areas for improvement, and discuss directions for future work.

2 Related Work

2.1 Relatedness in AI Understanding

Relatedness, alongside autonomy and competence, is one of the three core elements of self-determination theory for motivation and effective learning [28]. It refers to a sense of belonging, connectedness with others, and integration into social communities beyond oneself. Prior research has demonstrated that supporting relatedness enhances engagement and improves learning outcomes [29]. While many XAI tools aim to improve AI understanding for not just practitioners but non-technical users, studies find that individuals with limited technical backgrounds often lack relatable peers or role models, leaving them without a supportive network [39]. This absence of social connection reduces motivation and self-confidence, resulting in weaker learning outcomes. Similarly, Ehsan et al. [6] highlight the importance of social transparency, showing that observing how others in a community interact with AI systems can aid decision-making and promote understanding by creating a shared, social context for learning. These findings suggest that non-technical users face challenges in forming relatedness with XAI visualizations, limiting their ability to engage meaningfully. To address this gap, we aim to explore the use of LLM-driven narrative gamification to promote a sense of social connectedness, making the exploration of XAI visualizations more engaging and accessible for non-technical users.

2.2 Exploratory XAI Visualizations

Numerous exploratory XAI visualizations have been proposed to help users gain understanding of AI models. For instance, CNN Explainer [37] visualizes the neuron connections and pathways within a small CNN, allowing users to explore the interplay between its low-level mathematical operations and their high-level model structures. GAN Lab [16] promotes experimentation with Generative Adversarial Networks (GANs) by enabling users to interactively train them with a simple dataset. More examples of exploratory XAI visualizations include What-If Tool (WIT) [38] for model analysis in hypothetical situations, Manifold [41] for DNN comparisons, Squares [27] for multi-class model evaluation, RuleMatrix [23] for visualizing classifiers with rule-based knowledge representation, and DECE [5] for distilling ML models with counterfactual explanations.

Despite these visualizations proving effective for users with basic AI background, they are heavily techno-centric [6] and can be overwhelming for non-technical AI users. For instance, CNN Explainer [37] and GAN Lab [16], though advertised as suitable for non-experts, primarily visualize complex model structures like neuron activation pathways and layered distributions, concepts that are only comprehensible to those with basic ML understanding. Both tools also rely on technical terminologies (e.g., "discriminator," "generator," "conv_1_1,") to explain their visualizations, and used observational studies and simple usability questionnaires to demonstrate their effectiveness, lacking concrete study results on how much non-experts were able to "learn" through interaction. Similarly, WIT [38] was initially targeted at a broad audience including journalists and activists. However, after being tested with a variety of users, it became apparent that technical prerequisites were necessary to effectively explore the tool. Therefore, our study aims to explore the use of narrative gamification to provide context and guidance for individuals with limited AI background by personifying visualization elements into relatable game-like characters. By engaging users in a conversational manner with the visualization elements, we want to explore if narrative gamification can make the exploration experience more meaningful by creating an in-game support network, thus enhancing the experience and outcomes for non-technical users.

2.3 Narrative-based Gamification

In gamification, "narrative" involves incorporating storytelling elements and relatable characters to offer context and guidance outside of traditional game environments [10]. Narrative-based elements have been demonstrated to yield multiple positive effects, including enhancing learner memory, motivation, and engagement [3], as well as sparking interest in subjects otherwise deemed advanced or uninteresting [26]. For instance, Huynh et al. [13] introduced a role-playing game aimed at promoting visualization literacy in young children by leveraging the presence of narratives in data-related problems involving visualizations. Their study results showed that these elements improve engagement without sacrificing learning. Palomino et al. [25] developed and validated a Narrative Gamification Framework for Education, which provides educators with tangible guidelines to gamify their lessons by emphasizing the content's gameful transformation rather than the environment. The results of their user evaluations and expert feedback collectively demonstrated the effectiveness and potential of their framework in enhancing learner engagement, motivation, and learning outcomes in virtual learning environments.

Inspired by these works, we aim to explore narrative-based gamification within the context of XAI visualization, with the goal of enhancing non-technical AI users' understanding and engagement when exploring these tools. Specifically, our design probe introduces LLM-driven NPC-like visualization elements that are "context-aware" and capable of providing users with guidance on various aspects of the visualization. They are designed to address the lack of relatedness faced by non-experts when exporing XAI visualizations, which arises from the absence of a support network who can provide relatable guidance and help users navigate the complexities of XAI visualizations. Since gamification is known to effectively support relatedness [29], we explore the use of LLM-driven narrative gamification to create an in-game social support network. By transforming visualization elements into relatable game-like characters, we explore if gamification can help non-technical users better form social and emotional bonds, making their visualization exploration experience more meaningful and engaging.

3 Gamification Approach

Due to the complexity of XAI visualizations, non-technical users may find themselves puzzled by aspects like:

- **R1** Functionalities and usage of the interface view components;
- **R2** Technical terms and concepts mentioned in the visualization system that are relevant to AI/XAI;
- **R3** Meanings of visual encodings employed and how to interpret the actual visualizations.

In these cases, it is natural for them to want to seek assistance from a support network for learning and guidance. Our gamification "personifies" visualization elements into dynamic and relatable game-like characters to create this network. For this preliminary study, we chose to gamify interactive t-SNE projections of CNN embeddings because image classifiers are frequently encountered by non-technical AI users, such as in facial recognition apps [30] and home security cameras [32], as indicated in pre-questionnaires. Additionally, dimensionality reduction techniques like t-SNE are widely used in XAI visualization systems [11, 14, 15, 21]. Here, we present a scenario featuring our design probe of an interactive t-SNE projection of ResNet-34's CIFAR-10 embeddings to demonstrate the general approach of our gamification.

Suppose a non-technical user is exploring ResNet-34's model embeddings using our design probe. Upon launching the system, they notice that the interface is segmented into several views. The user finds the interface generally intuitive, except for the central view labeled as "*Projection View* (Figure 1-C)," which appears as a scatterplot filled with colorful points. Confused about its purpose and functionalities, the user clicks on a data point, instance #38, to initiate a conversation. A dialogue input textbox, along with the data point's avatar, depicted as a shy, blushing character, are displayed at the bottom. Instance #38 introduces itself, suggesting to the user, *"I-if you have any questions or need help understanding the projection, feel free to ask!"* The user inquires instance #38 by asking for an high-level explanation of the Projection View. Instance #38 answers, *"In the p-projection view, you can see where all the different d-data points are placed based on how s-similar they are to each other according to ResNet-34."* The user further asks for clarification on the view's buttons. Instance #38 responds, *"The buttons help you to z-zoom in, out, or reset the view its o-original position. You can also use the brush feature to select multiple data points for further exploration!"* Through this exchange and follow-up questions, the user gains an understanding of the view's purpose and features. This simplifies the initial learning phase, flattening the learning curve associated with adopting the tool (**R1**).

However, their exploration is still impeded due to their unfamiliarity with technical terms like "ResNet-34," "t-SNE," and "CIFAR-10." To familiarize with these AI/XAI terms, the user engages with instance #76 by initiating another conversation. Instance #76 is characterized as slightly annoyed by the user's presence, adding a unique dynamic to the learning experience. *"Ugh, what do you want now? [...] This t-SNE thingy shows our positions based on how similar we are. Got it?"* Confused about "t-SNE," the user asks for a brief overview of its concept. Instance #76 replies: *"t-SNE [...] is a technique used to visualize high-dimensional data in a lower-dimensional space, making it easier to see patterns and clusters in the data."* Through engaging in similar exchanges, the user develops an understanding of different technical terminologies used within the interface (**R2**).

Ready to explore the embeddings, the user wants to familiarize with the visual encodings. They observe two types of data points in the scatterplots: uniformly colored data points, and data points split into two halves. The user clicks on instance #79, depicted as a circle with its left half colored green and right half colored blue (Figure 1-1). Curious, the user asks the data point about the significance of its colors (Figure 1-2). Instance #79 explains, *"The colors represent the ground truth label on the left and ResNet34's prediction on the right. My green left half means I am a cat, while my blue right half means ResNet-34 predicted me as a Dog."* Knowing the color encoding, the user proceeds to examine the data distribution. They observe that ResNet-34 appears to struggle with inter-class discrepancy, particularly within the class "cat" (**R3**). They notice that ResNet-34 often

Figure 1: The interface of our design probe with LLM-powered narrative gamification. The components include: (a) *Overview View*, a summary of the dataset used and the model; (b) *Data Point(s) View*, which presents instance-level details of all selected data points; (c) *Projection View*, where users can freely explore and converse with the data points and clusters; (d) *Tasks & Notes View*, where users can record their tasks and collected insights, and (e) *Conversation History View*, where users can revisit their previous conversations with a data point.

misclassifies other subjects as cats and identifies several cat images as dogs. To understand this lack of fairness, the user activates the brush toggle button. The user drags their cursor to create a selection box around the dog cluster, thus highlighting every point predicted as a dog (Figure 1-4). A new avatar emerges, symbolizing the entirety of the dog cluster. The user begins by inquiring about the prediction accuracy, in which the cluster responds, *"In our cluster of 11 instances, eight were predicted correctly. The common misclassifications were instances of Cats being incorrectly predicted as Dogs."* The user then asks about the reasons for misclassification, to which the cluster responds: *"[...] It could be due to similarities in features like fur texture or body shapes that make them more challenging to differentiate."* This inspires the user to navigate to the Data Point(s) View (Figure 1-B) to examine the actual images within the cluster. They observe that the misclassified cat images are either of such low resolution, or feature cats with fur colors more typically associated with dogs. This similarity in visual features poses a challenge for ResNet-34 in accurately classifying these images. The user documents their insights regarding the model's frequent confusion between cats and dogs. As they gather more information from conversations, they record significant insights under their noted task, similar to how a player collects insights and takes notes while interacting with game NPCs (**R3**).

4 Preliminary Experiment and Results

To explore and evaluate how LLM-driven narrative gamification can help non-experts with explorations of XAI visualizations, we conducted a comparative study with participants possessing limited technical AI background and minimal familiarity with CNNs or t-SNE projections. Our preliminary investigation focuses on two aspects of the gamified approach: (**A1**) the effectiveness of gamification in improving users' understanding and insight gathering of XAI visualizations, and (**A2**) the impact of gamification on enhancing user engagement with the visualization tool.

4.1 Study Setup

Participants and Apparatus. We recruited 10 participants (P1 ∼ P10; six men, four women; age 22 ∼ 33) with limited technical background in AI/ML. They came from different backgrounds such as Civil Engineering, Computer Science, Industrial Engineering, and more. Specifically, on a 7-point Likert scale (self-rated; 1="Novice", 7="Expert"), we recruited participants that satisfied all the following constraints: AI experience ≤ 5, ML experience ≤ 4, familiarity with CNN models ≤ 2, and familiarity with t-SNE projections ≤ 2. These thresholds ensured that participants had limited exposure to AI/ML, so we could explore how gamification can aid non-technical AI users rather than those with extensive prior knowledge. Their median experience with AI is 2 (IQR=1), their median ML experience is 2 (IQR=0.75), their median CNN familiarity is 1 (IQR=1), and their median t-SNE familiarity is 1 (IQR=0). We also later verified their low prior knowledge of t-SNE and CNN through a pre-quiz assessing their understanding of these concepts. The study was conducted remotely via Zoom.

Task and Procedure. For our controlled experiment, we utilized a between-subjects study design, in which we compared our gamified prototype P_1 with a baseline variant P_0 that lacks the LLM-powered narrative gamification. Participants assigned to P_0 received an additional PDF document that included tutorials on the system interface and explanations of key AI/XAI concepts. The participants assigned to the baseline condition had an average AI experience of 2.8 (SD=1.303), and an average ML experience of 2 (SD=1.225). The participants assigned to the gamified condition had an average AI experience of 2.6 (SD=0.894), and an average ML experience of 1.8 (SD=0.837). We first required participants to take a pre-quiz consisting of eight questions to assess their general understanding of AI/XAI concepts. After introducing the assigned tool to the participants, they were provided with an online form requiring them to complete 12 tasks related to 1) understanding the system's interface, 2) grasping AI/XAI concepts, and 3) decoding visual encodings and interpreting the visualization. Participants had 40 minutes to complete these tasks and were encouraged to use the tasks & notes view (Figure 1-D) for documenting any notable insights about the visualization. To ensure authentic responses, participants were instructed to explain their answers in their own words, preventing direct copying and pasting of responses generated by LLMs. After the interaction, participants were asked to complete a Likert scale post-questionnaire that included NASA-TLX, which included six questions on the cognitive demand of task completion with their prototype. Additionally, there were 13 other questions that focused on the learning effects and usability of the system. Participants were then asked to take another post-quiz, which consists of 21 questions across three knowledge categories they were expected to learn during their interaction. To conclude each study session, an interview is conducted with each participant to gather additional qualitative data.

4.2 Results and Analysis: Task Performance

Quiz performance within groups. Through paired t-tests, we evaluated the learning gains within each condition, and found that both the baseline ($t = -3.763, p = 0.019$) and experimental groups ($t = -3.505, p = 0.024$) demonstrated statistically significant improvements. These results suggest that both systems are effective in enhancing users' understanding of AI/XAI concepts, regardless of the inclusion of the gamification feature.

Quiz performance and task completion between groups. In addition, we analyzed their average scores in quiz performance and task completion and utilized unpaired t-tests to statistically evaluate the differences between the groups. When examining the pre-quiz scores, participants in the gamified condition, on average, outperformed those in the baseline condition (M=68% > 60%). However, the unpaired t-test indicated that this difference is not statistically significant ($t = -0.524, p = 0.616$). Additionally, participants in the gamified condition achieved a high task completion rate of 100% (SD=0), compared to the baseline group's completion rate of 83.07% (SD=0.206). The unpaired t-test indicates that this difference is marginally statistically significant ($t = -0.1833, p = 0.141$). Despite the unpaired t-tests not revealing significant differences between the two groups, possibly due to the small sample size utilized by this study, participants from the gamified condition generally outperformed those in the baseline. This advantage was observed in task completion accuracy (M=84.61% > 75.38%; $t = -0.739, p = 0.498$), scores from the post-quiz questions on AI/XAI concepts (M=97.5% > 92.5%; $t = -0.600, p = 0.565$), and the overall post-quiz performance (M=91.43% > 88.57%; $t = -0.894, p = 0.406$).

4.3 Results and Analysis: Perceived Cognitive Load (NASA-TLX)

We employed the NASA-TLX to measure the perceived workload associated with each prototype. Both systems share similar mental (MD=5 for both conditions; p= 0.515) and physical demands (MD=2 for both conditions; p= 0.912), suggesting that the inclusion of gamification feature does not impose additional mental or physical strain on users. Compared to baseline, the gamified system has lower temporal demand (MD=5 < 6; p= 0.5219), leading to better performance (MD=3 < 5; p= 0.1363) and less effort (MD=4 < 5; p= 0.1931) with marginal statistical significance, as well as less frustration (MD=3 < 5; p= 0.6684). The overall perceived workload, measured by averaging all six raw NASA-TLX scores, was also lower for the gamified system than for the baseline (MD=3.833 < 4.667; p= 0.5296). Consequently, participants generally found the prototype integrated with our gamified framework to be less cognitively demanding compared to the baseline, although this difference was not statistically significant.

4.4 Results and Analysis: Perceived Learning Effects and Usability

Additionally, we asked the participants to self-rate their perceptions of the learning effects and usability of their assigned tool. Generally, participants from the gamified condition reported a better understanding of visualizations and AI/XAI concepts compared to those from the baseline. This was reflected in their perceptions of the amount of visualization understanding they gained (Q1: MD=6 > 4; p= 0.433), how much they felt they learned about CNN embeddings (Q4: MD=6 > 3; p= 0.110), and their perceived learning of t-SNE (Q5: MD=5 for both; p= 0.076), which exhibits a marginally significant p-value. This could be attributed to the gamification intuitively allowing users to seek explanations for concepts and visualizations that are unclear to them. For instance, while participants assigned to the baseline claimed that they could grasp the meanings of the positions and colors of the data points, they felt that the system offered only surface-level explanations an failed to provide in-depth insights into why certain data points were misclassified or positioned as they were within the projection. Participants assigned to the gamified condition commended the gamification for enhancing their understanding of the XAI visualization, attributing this improvement to the feature's ability to simplify the system's technical complexity. P4 mentioned, *"Using natural language made it easier for me to grasp and interact with the data points; I could learn further with follow-up questions for more detailed answers."* When examining ratings on interface understanding (Q3), no significant difference was observed between the two groups, as ratings from both were notably high (MD=7 for both conditions; p= 1.0). During interviews, participants explained that the prototype's interface was already straightforward and intuitive, and that those assigned to the gamified condition also did not find it necessary to rely on the gamification for assistance in navigating the interface.

Nonetheless, compared to the baseline, participants generally found the gamified tool to be more complex and difficult to use. This was reflected in their perceptions of whether the system was unnecessarily complex (Q6: MD=2 > 1; p= 0.287), its ease of use (Q7: MD=6 < 7; p= 0.083), and its ease of learning (Q9: MD=6 < 7; p= 0.193). Overall, the baseline system received higher ratings for its intuitiveness. Some participants from the gamified condition expressed that they felt it was unnecessary for users to have to hear visualization elements "speak" or to engage in conversations with them, as this disrupted their actual workflow while exploring the visualizations. Certain participants also found the gamified system more challenging to use due to the inherent inaccuracies of the GPT model used for the data points, as they noticed that sometimes these LLM agents provided answers that were clearly incorrect, leading to a loss of trust.

In terms of engagement, although the gamified prototype received a slightly higher rating than the baseline (Q10: MD=7 > 6; p= 0.908), participants assigned to the gamified group expressed mixed feelings. This ambivalence largely arises from the issues previously discussed, such as interruptions to the exploration workflow from needing to constantly interact with data points, frustration with the system due to inaccuracies (i.e., model hallucinations), and the excessive use of technical terms in explanations. Nonetheless, during the interviews, most participants still expressed positive sentiment towards the engagement of our gamification framework and highlighted the beneficial impacts it could have on non-technical users' exploration experience. P2 (gamified) explained, *"I love the feature of conversing with data points for its simplicity in obtaining answers and the charm of each point's unique personality, making the interaction engaging and informative."* P4 (gamified) stated, *"The system's gamified approach significantly boosts my motivation to study; it sparks my curiosity, encouraging me to explore further, revisit data points, and easily understand the classification through color differentiation, inspiring a continuous loop of discovery and learning."*

5 Discussion and Future Work

In discussing the use of narrative gamification for XAI visualizations, our study highlights certain key considerations. Firstly, our gamification has shown potential in motivating users to engage more deeply with the visualizations, sparking curiosity and encouraging a more proactive learning. This could also enhance the "replay" value of XAI visualizations, motivating users to revisit the tool for further exploration and uncover new insights. Secondly, another notable observation is that participants indeed experienced relatedness while interacting with the personified data points within our framework. Many participants shared during interviews that they felt they could ask any question without the fear of sounding uninformed or facing judgment. This demonstrates that our gamified approach can effectively create a relatable support network for non-experts to lower the barriers to AI learning. By promoting a more open and supportive exploration environment, our approach might encourage a wider audience of non-technical individuals to deepen their understanding of AI. However, a key concern that arises from our study is that some participants encountered hallucinated or incorrect AI-generated explanations, potentially misleading them and reducing their trust in the system.

For future work, we plan to conduct a formative study with target non-technical AI users to confirm the challenges identified in existing literature and explore additional obstacles that gamification may address. Furthermore, while this work serves as a preliminary investigation into the learning effects and user engagement of LLM-driven narrative gamification, we aim to conduct a more comprehensive study with a larger participant pool and explore gamification applied to other AI concepts and XAI visualization techniques. This expanded study will include detailed qualitative analysis of user workflows, observations, and interview feedback to answer further research questions, as well as additional conditions (e.g., guided explanations without NPCs) to better assess the specific impact of gamification. Future research will explore key questions, such as the pros and cons of combining gamification with XAI visualizations, the specific workflows users adopt when interacting with gamified XAI visualizations, and how users utilize gamification features to extract insights. Additionally, the impact of LLM hallucinations on user understanding and engagement could be further explored, along with strategies to mitigate them, such as prompting methods or retrieval-augmented generation (RAG) techniques. These findings will contribute to the development of empirical design insights, guiding future applications of gamification in XAI visualizations.

Acknowledgments

This work is supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant #RGPIN-2020-03966 and the University of Waterloo Interdisciplinary Trailblazer Grant.

References

[1] Inworld AI. 2023. Inworld Origins. https://store.steampowered.com/app/2199920/Inworld_Origins/ Available on Windows. Inworld AI..
[2] Nader Akoury, Qian Yang, and Mohit Iyyer. 2023. A Framework for Exploring Player Perceptions of LLM-Generated Dialogue in Commercial Video Games. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. 2295–2311. doi:10.18653/v1/2023.findings-emnlp.151
[3] Daniel R Bielenberg and Ted Carpenter-Smith. 1997. Efficacy of story in multimedia training. *Journal of Network and Computer Applications* 20, 2 (1997), 151–159. doi:10.1006/jnca.1997.0043
[4] Ilaria Caponetto, Jeffrey Earp, and Michela Ott. 2014. Gamification and education: A literature review. In *European Conference on Games Based Learning*, Vol. 1. Academic Conferences International Limited, 50.
[5] Furui Cheng, Yao Ming, and Huamin Qu. 2020. Dece: Decision explorer with counterfactual explanations for machine learning models. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 1438–1447. doi:10.1109/TVCG.2020.3030342
[6] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–19. doi:10.1145/3411764.3445188
[7] Raymond Fok, Nedim Lipka, Tong Sun, and Alexa F Siu. 2024. Marco: Supporting Business Document Workflows via Collection-Centric Information Foraging with Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20. doi:10.1145/3613904.3641969
[8] Laura Beth Fulton, Ja Young Lee, Qian Wang, Zhendong Yuan, Jessica Hammer, and Adam Perer. 2020. Getting playful with explainable AI: games with a purpose to improve human understanding of AI. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8. doi:10.1145/3334480.3382831

[9] Guodong Guo and Na Zhang. 2019. A survey on deep learning based face recognition. *Computer vision and image understanding* 189 (2019), 102805. doi:10.1016/j.cviu.2019.102805

[10] Stuart Hallifax, Maximilian Altmeyer, Kristina Kölln, Maria Rauschenberger, and Lennart E Nacke. 2023. From Points to Progression: A Scoping Review of Game Elements in Gamification Research with a Content Analysis of 280 Research Papers. *Proceedings of the ACM on Human-Computer Interaction* 7, CHI PLAY (2023), 748–768. doi:10.1145/3611048

[11] Fred Hohman, Haekyu Park, Caleb Robinson, and Duen Horng Polo Chau. 2019. Summit: Scaling deep learning interpretability by visualizing activation and attribution summarizations. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 1096–1106. doi:10.1109/TVCG.2019.2934659

[12] Wendy Hsin-Yuan Huang and Dilip Soman. 2013. Gamification of education. *Report Series: Behavioural Economics in Action* 29, 4 (2013), 37.

[13] Elaine Huynh, Angela Nyhout, Patricia Ganea, and Fanny Chevalier. 2020. Designing narrative-focused role-playing games for visualization literacy in young children. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 924–934. doi:10.1109/TVCG.2020.3030464

[14] Shichao Jia, Zeyu Li, Nuo Chen, and Jiawan Zhang. 2021. Towards visual explainable active learning for zero-shot classification. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 791–801. doi:10.1109/TVCG.2021.3114793

[15] Minsuk Kahng, Pierre Y Andrews, Aditya Kalro, and Duen Horng Chau. 2017. Activis: Visual exploration of industry-scale deep neural network models. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 88–97. doi:10.1109/TVCG.2017.2744718

[16] Minsuk Kahng and Duen Horng Chau. 2019. How does visualization help people learn deep learning? Evaluation of GAN Lab. In *IEEE VIS 2019 Workshop on EValuation of Interactive VisuAl Machine Learning Systems*. doi:10.1109/VIS47514.2020.00060

[17] Minsuk Kahng, Nikhil Thorat, Duen Horng Chau, Fernanda B Viégas, and Martin Wattenberg. 2018. Gan lab: Understanding complex deep generative models using interactive visual experimentation. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 310–320. doi:10.1109/TVCG.2018.2864500

[18] Sampo Kuutti, Richard Bowden, Yaochu Jin, Phil Barber, and Saber Fallah. 2020. A survey of deep learning applications to autonomous vehicle control. *IEEE Transactions on Intelligent Transportation Systems* 22, 2 (2020), 712–733. doi:10.1109/TITS.2019.2962338

[19] In Lee. 2021. Service robots: A systematic literature review. *Electronics* 10, 21 (2021), 2658. doi:10.3390/electronics10212658

[20] Tatiana Leshkevich and Anna Motozhanets. 2022. Social perception of artificial intelligence and digitization of cultural heritage: Russian context. *Applied sciences* 12, 5 (2022), 2712. doi:10.3390/app12052712

[21] Yuxin Ma and Ross Maciejewski. 2020. Visual analysis of class separations with locally linear segments. *IEEE Transactions on Visualization and Computer Graphics* 27, 1 (2020), 241–253. doi:10.1109/TVCG.2020.3011155

[22] Marek Milosz and Elzbieta Milosz. 2020. Gamification in engineering education–a preliminary literature review. In *2020 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 1975–1979. doi:10.1109/EDUCON45650.2020.9125108

[23] Yao Ming, Huamin Qu, and Enrico Bertini. 2018. Rulematrix: Visualizing and understanding classifiers with rules. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 342–352. doi:10.1109/TVCG.2018.2864812

[24] Zeinab Mokhtari, Ghasem Salimi, and Ali Akbar Safavi. 2021. Gamified Teaching-Learning Approaches in Higher Education: A scoping review. In *2021 14th National and 8th International Conference on e-Learning and e-Teaching (ICELET)*. IEEE, 1–8. doi:10.1109/ICELET202153025.2021.9712605

[25] Paula Toledo Palomino, Lennart Nacke, and Seiji Isotani. 2023. Gamification of virtual learning environments: a narrative and user experience approach. In *Anais do XXII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*. SBC. doi:10.1145/3638067.3638103

[26] Christos H Papadimitriou. 2003. MythematiCS: in praise of storytelling in the teaching of computer science and math. *ACM SIGCSE Bulletin* 35, 4 (2003), 7–9.

[27] Donghao Ren, Saleema Amershi, Bongshin Lee, Jina Suh, and Jason D Williams. 2016. Squares: Supporting interactive performance analysis for multiclass classifiers. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 61–70. doi:10.1109/TVCG.2016.2598828

[28] Richard M Ryan. 2017. *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.

[29] Michael Sailer, Jan Ulrich Hense, Sarah Katharina Mayr, and Heinz Mandl. 2017. How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in human behavior* 69 (2017), 371–380. doi:10.1016/j.chb.2016.12.033

[30] Sovantharith Seng, Mahdi Nasrullah Al-Ameen, and Matthew Wright. 2021. A first look into users' perceptions of facial recognition in the physical world. *Computers & Security* 105 (2021), 102227. doi:10.1016/j.cose.2021.102227

[31] Rita Sevastjanova, Wolfgang Jentner, Fabian Sperrle, Rebecca Kehlbeck, Jürgen Bernard, and Mennatallah El-Assady. 2021. Questioncomb: A gamification approach for the visual explanation of linguistic phenomena through interactive labeling. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 11, 3-4 (2021), 1–38. doi:10.1145/3429448

[32] Himani Sharma and Navdeep Kanwal. 2023. A Survey of Object Detection Techniques for Improving Smart Surveillance. In *2023 Seventh International Conference on Image Information Processing (ICIIP)*. IEEE, 603–609. doi:10.1109/ICIIP61524.2023.10537696

[33] Bumblebee Studios. 2023. Vaudeville. https://store.steampowered.com/app/2240920/Vaudeville/ Available on Windows and Linux. BBS Games AB..

[34] Shankar Sundaresan and Zuopeng Zhang. 2022. AI-enabled knowledge sharing and learning: redesigning roles and processes. *International journal of organizational analysis* 30, 4 (2022), 983–999.

[35] TensorFlow Team. 2024. TensorFlow Embedding Projector. https://projector.tensorflow.org/. Accessed: 2024-04-07.

[36] Amrita S Tulshan and Sudhir Namdeorao Dhage. 2019. Survey on virtual assistant: Google assistant, siri, cortana, alexa. In *Advances in Signal Processing and Intelligent Recognition Systems: 4th International Symposium SIRS 2018, Bangalore, India, September 19–22, 2018, Revised Selected Papers 4*. Springer, 190–201. doi:10.1007/978-981-13-5758-9_17

[37] 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. doi:10.1109/TVCG.2020.3030418

[38] James Wexler, Mahima Pushkarna, Tolga Bolukbasi, Martin Wattenberg, Fernanda Viégas, and Jimbo Wilson. 2019. The what-if tool: Interactive probing of machine learning models. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 56–65. doi:10.1109/TVCG.2019.2934619

[39] Qi Xia, Thomas KF Chiu, Min Lee, Ismaila Temitayo Sanusi, Yun Dai, and Ching Sing Chai. 2022. A self-determination theory (SDT) design approach for inclusive and diverse artificial intelligence (AI) education. *Computers & Education* 189 (2022), 104582. doi:10.1016/j.compedu.2022.104582

[40] Yuzhe You, Jarvis Tse, and Jian Zhao. 2023. Panda or not Panda? Understanding Adversarial Attacks with Interactive Visualization. *arXiv preprint arXiv:2311.13656* (2023).

[41] Jiawei Zhang, Yang Wang, Piero Molino, Lezhi Li, and David S Ebert. 2018. Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 364–373. doi:10.1109/TVCG.2018.2864499