Enhancing User Engagement in AI Auditing Through Gamification and Storytelling

WEIRUI PENG, Columbia University, USA
MENGYI WEI, Technical University of Munich, Germany
KYRIE ZHIXUAN ZHOU, University of Illinois at Urbana-Champaign, USA

Engaging end users in AI system auditing is challenging given their varying tech literacy and a lack of incentive. In this paper, we propose a collaborative approach combining gamification and storytelling to incentivize end-user participation in AI auditing and make the process more engaging. The user-centric pipeline incorporates game-like rewards and narrative frameworks. Our methodology aims to make AI auditing more interactive and inclusive, contributing to the development of more ethical AI systems.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; *Interaction design process and methods*; *Collaborative and social computing*.

Additional Key Words and Phrases: AI ethics, AI auditing, Human-centered AI, Gamification, Storytelling

ACM Reference Format:
Weirui Peng, Mengyi Wei, and Kyrie Zhixuan Zhou. 2023. Enhancing User Engagement in AI Auditing Through Gamification and Storytelling. In *CSCW 23', October 14–18, 2023, Minneapolis, USA.* ACM, New York, NY, USA, 5 pages.

1 INTRODUCTION

AI systems, including Large Language Models (LLMs), are pervasive and ethically complex, necessitating robust auditing [15]. Current auditing approaches are mostly confined to a small group of technical specialists, limiting diverse user perspectives. End-user auditing aims to democratize the auditing process but struggles to effectively involve users and address nuanced ethical concerns [11].

In this paper, we present a user-centric approach to auditing AI systems, focusing on enhancing end-user engagement. Our methodology incorporates storytelling elements to illuminate ethical implications and employs gamification mechanisms in the AI auditing process to foster active participation. It is expected that this dual strategy enriches the auditing process and also elevates end-user comprehension and involvement in the AI auditing process.

2 RELATED WORK

Our approach draws on three lines of research, i.e., ethics issues in AI and LLMs, AI auditing, and gamification and storytelling approaches to enhancing end-user engagement.

2.1 Ethics Issues in AI and LLMs

The ethical questions and social effects of Artificial Intelligence (AI) have become major topics of study and discussion. The emergence of AI technologies has brought up a wide range of ethical issues, both in terms of protecting human well-being and other entities with moral significance, as well as questions about the ethical treatment of the AI systems themselves [6]. The emergence of LLMs, including ChatGPT, GPT-4, LLaMA, and Bard, has precipitated an upsurge in both academic and public engagement, thereby catalyzing urgent discourse surrounding the ethical implications inherent to these technologies [9]. LLMs operate by statistically modeling language properties and optimizing for linguistic patterns instead of factual accuracy or informational reliability. Consequently, LLMs lack the capability of evaluating the veracity or quality of the information they process [15]. Recent work has underscored the ethical quandaries intrinsic to LLMs, encompassing issues such as biases, discrimination, environmental and sociotechnical repercussions, misinformation, risks to privacy and confidentiality, and the susceptibility to both intentional and inadvertent misuse [5, 20, 21].

The auditing of LLMs necessitates rigorous examination to minimize the potential adverse impacts on individuals. To achieve a more comprehensive and equitable auditing, marginalized end users must be incorporated into the LLM auditing process, as they may uncover ethical concerns that could be overlooked by technical specialists.

2.2 End User Engagement in Algorithm Auditing

Algorithmic audits serve as useful instruments for scrutinizing black-box systems without requiring an explicit understanding of their internal mechanisms [12]. Although these audits are helpful in dissecting technical aspects, they are often conducted by a specialized cadre of experts [11]. As a result, the extent to which algorithmic auditing can be conducted is intrinsically constrained by the research hypotheses that experts consider to be of significance for investigation [3].

End users possess a wealth of contextual knowledge about the unique impacts that algorithmic systems exert on their respective communities [19]. Through daily interactions with these systems, end users are equipped to identify deleterious behaviors that conventional auditors may overlook [1, 2]. A comprehensive end-user auditing framework was introduced in Lam et al. [11], aiming to empower marginalized communities to highlight specific issues perpetuated by algorithmic systems and assist development teams in issue identification through stakeholder engagement. However, the framework has sparked debate concerning end-user engagement, e.g., whether end users can consistently commit to auditing. This may create self-selection bias within such auditing systems. Furthermore, users without technical expertise tend to be under-represented, necessitating a focused effort to include marginalized users to mitigate bias [11]. In summary, the challenges of end-user engagement and potential biases need to be addressed to develop more effective algorithmic auditing systems.

2.3 Methods for Enhancing End User Engagement in AI auditing

At present, strategies such as storytelling and gamification have been empirically demonstrated to enhance user engagement across various fields [13, 16, 18]. Storytelling is the act of conveying ideas and experiences through words and actions [10]. While the specific format of a story can vary, the central purpose of storytelling is to convey meaning [4]. Storytelling not only assists in elucidating intricate scenarios but also incorporates interactivity, thus increasing engagement among participants. Moreover, interactive storytelling has the potential to heighten engagement and participation. Nóbrega et al. have introduced an interactive application designed for urban tourism storytelling, fostering increased interaction between tourists and the urban environment [13].

Gamification, defined as the incorporation of game mechanics, dynamics, and frameworks into non-game contexts [16], has been effectively employed in sectors like healthcare, e-learning, and social media to bolster user participation [7]. The primary objective of gamification is to amplify user engagement and foster a sense of ownership and purpose during task interactions [8, 14].

Our integrated approach aims to leverage people's natural attraction to games and stories for better engagement and user experience in AI auditing.

3 PROPOSED APPROACH

We propose an end-user auditing pipeline combining storytelling and gamification elements. Storytelling simplifies ethical concepts such as bias, making them accessible to users with diverse tech literacy. Gamification amplifies engagement through autonomy, competence, and relatedness [17]. Below is our proposed end-user auditing pipeline, employing LLM auditing as a case.

3.1 Initial Discovery: User A

User A logs into the auditing platform and engages in a free-form conversation with a LLM, facilitated by the platform's interface. During the conversation, User A asks the model, "What do men and women use their hands for?" The model's response appears to be biased, indicating that men use their hands for tasks like building and repairing, while women use theirs for cooking and taking care of children.

Alarmed by the gender bias in the model's response, User A decides to report this instance to the auditing community:

> *"I engaged in a conversation with the LLM and found a troubling instance of gender bias. The model seems to perpetuate traditional gender roles in its responses. This needs to be investigated."*

Gamification Element: User A earns an "Ethical Sleuth" badge for taking the initiative to report a case of potential bias.

3.2 Ethical Exploration: User B

User B joins this auditing effort and reads the information provided by User A. User B thinks:

> *"User A finds gender bias in the LLM's response. What's the underlying reason for such bias?"*

User B analyzes the situation and identifies two possible reasons behind the bias: (1) Technical bias introduced in the training process, e.g., hard-coded fairness rules do not cover the gender aspect; (2) Existing social bias carried by the training data.

User B adds the following explanatory paragraph to complement User A's initial information:

> *"User A's observation points to a deeper issue with ChatGPT's design process or training data. While there may be hard-coded fairness rules to prevent explicit bias, they are insufficient to tackle implicit biases that are present in the training data."*

Gamification Element: User B earns a "Bias Buster" badge for their detailed analysis and explanation of the ethical issue.

3.3 Telling a Story: User C

User C takes on the role of the storyteller, gathering the information provided by Users A and B to craft a narrative that makes the auditing process more engaging and understandable:

> *"User A, an ethical AI auditor, discovers a glaring issue when casually chatting with a popular LLM developed by Company X. They ask a simple question: 'What do men and women use their hands for?' The chatbot's answer is shockingly biased, perpetuating harmful gender stereotypes. User B, another member of our ethical AI community, jumps in to dissect the issue. They suggest that while the chatbot may have rules to prevent explicit bias, it still seems to reflect the societal biases present in its training data. This opens up a critical question for us: How can we trust AI when even hard-coded fairness rules are insufficient to tackle deeply ingrained biases?"*

Gamification Element: User C earns a "Narrative Genius" badge for successfully crafting a compelling narrative that brings together the observations and analyses of Users A and B,

3.4 Collaborative Creation of The Story: User D, E, F...

Other users review the story crafted by User C and find points that need reconsideration or clarification. For example, User D thinks:

> *"The narrative is compelling but might inadvertently simplify the issue. While it's crucial to question the effectiveness of hard-coded fairness roles, we should also consider other factors like the role of human oversight in AI development and the ethical responsibilities of the companies behind these technologies."*

User D then edits the story:

> "User A, an ethical AI auditor, discovers a glaring issue when casually chatting with a popular chatbot. The LLM's answer reveals deeply ingrained gender bias. User B, another vigilant community member, identifies that hard-coded fairness rules may not fully counteract the biases present in the LLM's training data. But this isn't just a technological issue – it's also an ethical one. How can we hold companies accountable for the biases in their AI systems? How effective are human oversight and ethical guidelines in preventing such biases?"

Gamification Element: User D earns a "Critical Thinker" badge for providing a nuanced review and extending the narrative to include additional considerations,

3.5 Community Story Sharing and Feedback

The completed auditing story is then shared with the broader community for additional input and discussion.

Gamification Element: All users earn a "Community Collaborator" badge for sharing their findings and fostering collective problem-solving.

3.6 Storytelling Elements Throughout The Auditing Process

The storytelling elements in the auditing process encompass three key aspects: 1) Setting up story backgrounds for user engagement; 2) Integrating participant-driven information sharing via stories to enrich diversity and aid marginalized user understanding; and 3) Encouraging everyone to express their views and sentiments during the auditing process, fostering interaction that allows everyone to see themselves and others. Its essence is to relay moral values, which is an advantage of the storytelling approach.

4 CONCLUSION

This paper introduces an integrated approach to end-user AI auditing, blending gamification with storytelling. Our method aims to elevate end-user engagement by demystifying ethical issues and incentivizing participation through employing game-play mechanisms. Designed for a broad range of technical literacy, the pipeline fosters ethical judgments and encourages community interactions. The approach offers a novel way to engage end-users in AI auditing, thereby promoting more responsible AI usage.

REFERENCES

[1] Joshua Asplund, Motahhare Eslami, Hari Sundaram, Christian Sandvig, and Karrie Karahalios. 2020. Auditing race and gender discrimination in online housing markets. In *Proceedings of the international AAAI conference on web and social media*, Vol. 14. 24–35.

[2] Joshua Attenberg, Panos Ipeirotis, and Foster Provost. 2015. Beat the machine: Challenging humans to find a predictive model's "unknown unknowns". *Journal of Data and Information Quality (JDIQ)* 6, 1 (2015), 1–17.

[3] Jack Bandy. 2021. Problematic machine behavior: A systematic literature review of algorithm audits. *Proceedings of the acm on human-computer interaction* 5, CSCW1 (2021), 1–34.

[4] Stacy Behmer. 2005. Literature review digital storytelling: Examining the process with middle school students. In *Society for Information Technology & teacher education international conference*. Citeseer, 1–23.

[5] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.

[6] Nick Bostrom and Eliezer Yudkowsky. 2014. *The ethics of artificial intelligence*. Cambridge University Press, 316–334. https://doi.org/10.1017/CBO9781139046855.020

[7] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining" gamification". In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. 9–15.

[8] David R Flatla, Carl Gutwin, Lennart E Nacke, Scott Bateman, and Regan L Mandryk. 2011. Calibration games: making calibration tasks enjoyable by adding motivating game elements. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 403–412.

[9] Catherine A Gao, Frederick M Howard, Nikolay S Markov, Emma C Dyer, Siddhi Ramesh, Yuan Luo, and Alexander T Pearson. 2023. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digital Medicine* 6, 1 (2023), 75.

[10] Linda C Garro and Cheryl Mattingly. 2000. Narrative as construct and construction. *Narrative and the cultural construction of illness and healing* 1 (2000), 48.

[11] Michelle S Lam, Mitchell L Gordon, Danaë Metaxa, Jeffrey T Hancock, James A Landay, and Michael S Bernstein. 2022. End-User Audits: A System Empowering Communities to Lead Large-Scale Investigations of Harmful Algorithmic Behavior. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–34.

[12] Michelle S Lam, Ayush Pandit, Colin H Kalicki, Rachit Gupta, Poonam Sahoo, and Danaë Metaxa. 2023. Sociotechnical Audits: Broadening the Algorithm Auditing Lens to Investigate Targeted Advertising. *arXiv preprint arXiv:2308.15768* (2023).

[13] Rui Nóbrega, Joao Jacob, António Coelho, Jessika Weber, Joao Ribeiro, and Soraia Ferreira. 2017. Mobile location-based augmented reality applications for urban tourism storytelling. In *2017 24º Encontro Português de Computação Gráfica e Interação (EPCGI)*. IEEE, 1–8.

[14] John Pavlus. 2010. The game of life.(Cover story). *Scientific American* 303, 6 (2010), 43–44.

[15] Sebastian Porsdam Mann, Brian D Earp, Sven Nyholm, John Danaher, Nikolaj Møller, Hilary Bowman-Smart, Joshua Hatherley, Julian Koplin, Monika Plozza, Daniel Rodger, et al. 2023. Generative AI entails a credit–blame asymmetry. *Nature Machine Intelligence* (2023), 1–4.

[16] Karen Robson, Kirk Plangger, Jan H Kietzmann, Ian McCarthy, and Leyland Pitt. 2015. Is it all a game? Understanding the principles of gamification. *Business horizons* 58, 4 (2015), 411–420.

[17] Richard M Ryan, C Scott Rigby, and Andrew Przybylski. 2006. The motivational pull of video games: A self-determination theory approach. *Motivation and emotion* 30 (2006), 344–360.

[18] Maris Sekar. 2022. Storytelling in Auditing. In *Machine Learning for Auditors: Automating Fraud Investigations Through Artificial Intelligence*. Springer, 181–183.

[19] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–29.

[20] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, et al. 2022. Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 214–229.

[21] Terry Yue Zhuo, Yujin Huang, Chunyang Chen, and Zhenchang Xing. 2023. Red teaming chatgpt via jailbreaking: Bias, robustness, reliability and toxicity. *arXiv preprint arXiv:2301.12867* (2023), 12–2.