User experience design of AI-assisted human-technology ecosystem for writing assessment

Liang, L.

academic integrity
Generative AI
GPTZERO
Human IA (intelligence augmentation)
User Experience Design
Writing Assessment

Generative AI (GenAI) raises increasing concerns as it has been swiftly used for academic misconduct and the tech communities have not figured out effective solution to detect AI-written text, particularly in the context of writing assessment. This threat to academic integrity and assessment validity urgently calls for effective change in the user experience (UX) design of writing assessment before we can detect AI-written text one day. How to build up a sustainable and valid GenAI-assisted human-technology ecosystem for writing assessment with academic integrity? What practical goals of writing assessment can we use GenAI agents to meet and extend to its full potential for human IA (intelligence augmentation)? This experimental pilot study reviews the user experience design of four GenAI tools and GPTZERO with socio-cultural approach to explore these questions from user-centered perspective. Three principles for user experience design of AI-assisted writing assessment are synthesized for mastery learning.

Introduction

Almost every time when a record-breaking new technology came into being in the human history, be it electricity over candles, or cars over horses, we users were confronted with the similar process of shock, doubt, rejection, fight, taste, confirmation, accept, like, prevalent vogue everywhere, and so on. The similar story happens with generative AI (GenAI) tools such as ChatGPT, Claude, Grok and Gemini. On the other hand, GenAI raises increasing concerns as it has been swiftly used for academic misconduct and the tech communities have not yet figured out effective solution to detect AI-written text, particularly in the context of writing evaluation or assessment. For example, the International Conference on Machine Learning (ICML), has officially forbid researchers from using AI tools like ChatGPT to write scientific papers. Though GPTZERO can detect most of AI-written work, it cannot when some words in the AI-written work are changed by humans.

Given such facts mentioned above, how can we effectively and validly design our user experience (UX) of GenAI-assisted human-technology ecosystem for the writing assessment? What practical goals of writing assessment can we use GenAI agents to meet and extend to its full potential for human IA (intelligence augmentation)?

Methods

This experimental pilot study conducts a small-scale experiment with a sample of four GenAI tools--- ChatGPT, Claude, Grok and Gemini, and one AI-writing detection tool GPTZERO, to explore and review the UX design of using GenAI tools and GPTZERO in writing assessment with socio-cultural approach. The sampling details, prompts, interaction logs as empirical evidence from which key conceptual principles were derived in this paper are illustrated in GitHub repository https://github.com/llcfy2023/User-experience-design-of-GenAI-assisted-human-technology-ecosystem-for-writing-assessment/tree/main. Content analysis in socio-cultural study was adopted for the data analysis.

Results

Based on the human-centered nested learning engineering cycle (Goodell, Kessler & Schatz, 2023; Craig et al., 2025), socio-cultural theory (Vygotsky, 1986; Lave & Wenger, 1991) and modern assessment theory (Andrich & Marais, 2019; Tognolini & Stanley, 2007), three principles for user experience design for AI-assisted writing assessment are synthesized for UX design of mastery learning: socio-cultural fitness, incremental validity, and empower human IA for all people. A new writing assessment mode of purpose is then proposed for the front-line educators in GenAI age.

Discussion

1. socio-cultural fitness

Undeniably, the digital technology like internet and GenAI not only expands our knowledge/information to a historically unprecedented level, but also brings us new challenges and enhances our ability to do things at an unheard-of new level ever. We cannot ban internet from our students' life in the internet age, nor can we forbid their use of GenAI in AI age. Mathematician Dr. Terence Tao in UCLA noted in his 2022 mathstodon post that it is of no use to fight against the new tech in the long term and educators need to adopt an "open books, open AI" mode of assessment to adapt to the times.

The search engines like google can offer us information on our questions, but what makes GenAI more intelligent than the search engines are that it fits more with human culture in intelligent interactions and enquiries, not just at the interactional level of accessing knowledge or information as google, but also at the interactional level of more or less synthesized ideas or thinking. This design path based on socio-cultural fitness is obviously different from the conventional path based on neuro-mimic design. The better fit with the human culture of interaction that AI tools can have, the more human-like intelligent and instant popularized they will be. An AI tool that fits well with human culture and can provide us the same assistant service that we need for our life in human society, can be another successful way besides the neuro-mimic AI development to realize AI design to its record-breaking level.

With the nature of socio-cultural fitness, the goal of UX design like ChatGPT and GPTZERO is to empower the human culture of interaction and communication in learning as a collaborative intelligent assistant in response to the people's prompts, not to replace people in the division of labor in the digital activity of GenAI age. The principle of socio-cultural fitness also offers UX design in GenAI an effective and simple means to avoid ethical pitfalls that may potentially threaten or harm the healthy cultural norms of the human society.

Another outstanding benefit of this socio-cultural fitness in UX design is that it can greatly enhance the UX value of usability and desirability. How long will it take you to learn to use ChatGPT or GPTZERO? Often only within a couple of seconds. This is because the system interacts just like a real person, providing us a genuine natural human communication experience of a chat that we have already been very familiar with.

1. incremental validity

Incremental validity is the second key principle for effective UX design of GenAI tools in the context of writing tasks. Its natural dialogue to exchange ideas fits well for providing feedback, answering questions in formative assessment (FA) as intelligent agents in learning and assessment. But a major reason for the low acceptability in some cases of evaluation or even a ban on ChatGPT in New York public schools is that the UX design of ChatGPT omitted any technical measures to increase the evidence-based validity of its answers in the users' evaluation, nor provide specific sources of reference to the users for more trustworthy human-technology interaction and collaboration. In contrast, the UX design of GPTZERO pinpoints exactly this principle of incremental validity: GPTZERO clarifies the evidence-based measures and evaluative framework that the intelligent analysis outcome is based on. This undoubtedly enhances the transparency of its background operation and increases the trustworthiness in its users.

2. Empower human IA for all learners via GenAI

If AI tools are designed for the coding people only, then that means a majority of the population will be left out, which is in opposite to our goal for an inclusive and sustainable learning engineering. This principle serves to structure the AI-assisted human-technology ecosystem as a user-centered inclusive and sustainable one for learning and assessment. The change unavoidably leads to a change in what values in the workforce market: those who are excellent in collaboration with AI tools to solve problems and complete tasks with good efficiency at work will be in great demand in the near future. This new change in the values of collaborative capacity with AI tools in the workforce market will of course prompt the need for a new assessment mode for the students using AI agents, especially for the writing assessment.

3. A new writing assessment mode of purpose

This new mode will be discussed via two key questions: what to assess and how to assess. Before figuring out these two questions, we'll need to be clear of the new purpose of the assessment mode for the students using AI tools. As discussed above, the purpose of the new assessment mode is not to forbid the students from using AI tools, but to assess how well the students complete the assessment task in collaboration with AI tools such as ChatGPT and GPTZERO.

Firstly, we will have to redefine what to assess in the writing assessment. With the open access to internet and availability of AI tools to all students such as ChatGPT, it is of somewhat limited value to assess the students' knowledge and basic thinking skills such as memory anymore, since everyone can get equal access to internet for all human knowledge over the thousands of years in a second. What matters more than merely knowledge and basic thinking skills in the digital age, is the students' ability in problem-solving, higher-order thinking, collaboration with other students and AI tools, value of the culture and society, etc.

According to the structure of the activity theoretical framework (Leontiev, 1978), it is also necessary to consider the new change in the division of labor in the digital activity of learning: the students will shoulder more responsibility as independent leaders, managers or principal investigators in their collaboration with AI tools, different from their traditional role of somewhat passive object to follow the teachers' leadership. Thus, the students' leadership and organization ability in their collaboration with AI tools in accomplishing the assessment tasks may also be included in the content of the writing assessment in university.

Another key question is how to assess the students using AI tools in the context of writing assessment in university. The teachers who would like to give their students writing assessment tasks will need to refresh their repertoire of assessment means and strategies according to their specific assessment purposes. AI-assisted human-technology ecosystem works effectively where speed, encyclopedic knowledge, or precision are key for a writing assessment task. For example, GenAI provides a short-cut to the students who need to complete a writing task but donot have sufficient time. If the teachers want to assess how well the students can collaborate with GenAI tools to complete a writing assessment task in a short time, then the teachers may design a writing assessment task that requires the students to finish in collaboration with AI tools within a very short period of time. Last but not least, teachers also need to be careful in their wording in the writing assessment task description so as to establish the identity of AI tech as assistant tools only, not as another robotic agent beings for the task.

References

ChatGPT Homepage, https://chat.openai.com/auth/login, last accessed 2025/10/15.

Claude Homepage, https://claude.ai/chat, last accessed 2025/10/15.

Gemini Homepage, https://gemini.google.com/app, last accessed 2025/10/15.

GPTZERO Homepage, http://www.GPTZERO [2].me, last accessed 2025/10/15.

Grok Homepage, https://www.grok.com, last accessed 2025/10/15.

Andrich, D., & Marais, I. (2019). A course in Rasch measurement theory: Measuring in the educational, social and health sciences. Springer Singapore.

Craig, S. D., Avancha, K., Malhotra, P., C., J., Verma, V., Likamwa, R., Gary, K., Spain, R., & Goldberg, B. (2025). Using a Nested Learning Engineering Methodology to Develop a Team Dynamic Measurement Framework for a Virtual Training Environment. In International Consortium for Innovation and Collaboration in Learning Engineering (ICICLE) 2024 Conference Proceedings: Solving for Complexity at Scale (pp. 115-132). https://doi.org/10.59668/2109.21735

Goodell, J., Kessler, A., & Schatz, S. (2023). Learning Engineering at a Glance. Journal of Military Learning, https://www.armyupress.army.mil/Journals/Journal-of-Military-Learning/Journal-of-Military-Learning-Archives/Conference-Edition-2023-Journal-of-Military-Learning/Engineering-at-a-Glance/

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge University Press.

Tao, T. (2022). AI tools like #ChatGPT will soon be capable of answering a large fraction of traditional university homework-type questions [Social media post]. Mathstodon. https://mathstodon.xyz/@tao/109543141003492779

Tognolini, J., & Stanley, G. (2007). Standards-based assessment: a tool and means to the development of human capital and capacity building in education. The Australian Journal of Education, 51(2), 129-145. https://doi.org/10.1177/000494410705100203

Vygotsky, L. S. (1986). Thought and language. MIT Press.