

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

Battle of the Bots: Generative AI tools head-to-head

Penny Wheeler

University of Technology Sydney

Colin Simpson

Deakin University

Wendy Taleo

Flinders University

Antony Tibbs

Edith Cowan University

The ongoing expansion of Generative Artificial Intelligence (GenAI) technologies since the launch of OpenAI's ChatGPT in late 2022 has created an unprecedented level of uncertainty about learning and teaching practice in the tertiary education sector. The rapid pace of change and the steady release of new tools and functionalities have made staying informed a challenge for time-poor educators, students, institutional leaders and third space practitioners. This light-hearted panel discussion aims to alleviate some of this stress by showcasing the strengths and weaknesses of some of the leading GenAI platforms by comparing their performance when tasked with a selection of learning and teaching-related use cases. Our goal for the session is to provide attendees with an opportunity to discuss the important characteristics of useful GenAI applications as we test and develop an assessment rubric for the output from the bots and evaluate the tools together.

Keywords: GenAI, evaluation, human-AI, curriculum design, learning support

Area of focus

The Battle of the Bots explores the functionality of several generative AI (GenAI) agents used for student learning and support by comparing their responses for tasks drawn from educational scenarios. This session will consider how the choice of GenAI platform influences the quality of learning and teaching activities in specific circumstances. It will also foster discussion about what qualities of these tools are of greatest importance to educators, students and other institutional stakeholders.

Within the conference subtheme of "AI and Human Synergy", and the conference's dynamic landscape that is the now and the future, we seek to bring the audience into a debate on the relative value of the operation and outputs of GenAI agents or 'bots'.

Why this approach is of interest

Much of the discussion about the application of GenAI tools in learning and teaching centres on examples using single individual platforms. A comparative approach creates opportunities to consider these technologies more deeply, an approach which we (and UC Berkeley, n.d.) have termed a "battle", but which has also been recommended and presented as an AI Competition (Sculley et al., 2025).

Theoretical frameworks

As we are in an era where GenAI technologies are disrupting teaching and learning and making it necessary for new ways of thinking (Bozkurt, 2023), this debate is critical as well as seriously playful (Rieber et al., 1998). This session will be informed by principles of realist evaluation, "a branch of theory-based evaluation specifically for the evaluation of complex social interventions ... A realist approach assumes that nothing works

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

everywhere for everyone and that context really makes a difference” (King et al., 2016, pp. 21). Realist evaluation has been increasingly used in the evaluation of educational technologies in recent years (Bryant et al., 2019) to accommodate the complex environments in which they are used.

The starting point for the draft rubric to compare the outputs from the different tools is the conceptual framework designed specifically for evaluating GenAI conversational agents developed by Yusuf et al. (2025). Discussion points in the session might arise around the parameters of “good AI” (Dash, 2025) and the impact of GenAI on student learning experience (Fawns et al., 2024), and these are likely to extend the rubric further.

Panellists and audience

We welcome educators, institutional leaders and third space practitioners with an interest in GenAI to be part of the audience, and to serve as the bot evaluators, for this session.

Panellists from the Australasian higher education community of GenAI specialists have been asked to develop chatbots in their preferred platform (different for each panellist) ready for the battle. During the session, the panellists will be tasked with demonstrating the performance of these chatbots in response to challenge activities selected by the audience, using resources (for example, subject outlines, sample assignment submissions, or learning analytics data) supplied by the organisers and previously unseen.

Attendees will use the draft assessment rubric to evaluate the competing bots and share suggestions for improvement and application, thereby gaining further practical understanding of the affordances and value of GenAI platforms used in education.

References

- Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, 18(1).
<http://www.asianjde.com/ojs/index.php/AsianJDE/article/view/718>
- Bryant, M., Ryan, K., Simpson, C., Whitelock-Wainwright, A., & Wood, T. (2019). At the heart of a diverse technology: Applying a realist evaluation methodology to a university live streaming programme. In S. C. Y. Wei, C. K. Mun, & A. Alphonso (Eds.), *Personalised learning. Diverse goals. One heart. ASCILITE 2019* (pp. 59–67), Singapore. <https://2019conference.ascilite.org/assets/proceedings/ASCILITE-2019-Proceedings-Final.pdf>
- Dash, A. (2025, May 1). *What would “good” AI look like?* Anil Dash.
<https://www.anildash.com/2025/05/02/what-would-good-ai-look-like/>
- Fawns, T., Henderson, M., Matthews, K., Oberg, G., Liang, Y., Walton, J., Corbin, T., Bearman, M., Shum, S. B., McCluskey, T., McLean, J., Shibani, A., Bakharia, A., Lim, L.-A., Pepperell, N., Slade, C., Chung, J., & Seligmann, A. (2024). Gen AI and student perspectives of use and ambiguity: A multi-institutional study. *ASCILITE Publications*, 132–134. <https://doi.org/10.14742/apubs.2024.1218>
- King, M., Rothberg, S., Dawson, R., & Batmaz, F. (2016). Bridging the edtech evidence gap: A realist evaluation framework refined for complex technology initiatives. *Journal of Systems and Information Technology*, 18(1). <https://doi.org/10.1108/JSIT-06-2015-0059>
- Rieber, L. P., Smith, L., & Noah, D. (1998). The value of serious play. *Educational Technology*, 38(6), 29-37.
- Sculley, D., Cukierski, W., Culliton, P., Dane, S., Demkin, M., Holbrook, R., Howard, A., Mooney, P., Reade, W., Risdal, M., & Keating, N. (2025). *Position: AI competitions provide the gold standard for empirical rigor in GenAI evaluation* (No. arXiv:2505.00612). arXiv. <https://doi.org/10.48550/arXiv.2505.00612>
- UC Berkeley. (n.d.) *LMarena* [Website]. <https://lmarena.ai/>
- Yusuf, H., Money, A., & Daylamani-Zad, D. (2025). Pedagogical AI conversational agents in higher education: A conceptual framework and survey of the state of the art. *Educational Technology Research and Development*, 73(2), 815–874. <https://doi.org/10.1007/s11423-025-10447-4>

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

Wheeler, P., Simpson, C., Taleo, W., & Tibbs, A. (2025, Nov 30 – Dec 3). Battle of the Bots: Generative AI tools head-to-head [Panel Presentation]. Australasian Society for Computers in Learning in Tertiary Education Conference, Adelaide, Australia. <https://doi.org/10.65106/apubs.2025.2778>

Note: All published papers are refereed, having undergone a double-blind peer-review process.

The author(s) assign a Creative Commons by attribution licence enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.

© Wheeler, P., Simpson, C., Taleo, W., & Tibbs, A. 2025