

Evaluating and transforming you and your colleagues' assessments in an age of generative AI

P. R. Neal¹ School of Chemical Engineering, University of New South Wales Sydney, Australia https://orcid.org/0000-0002-8831-5327

S. Daniel School of Professional Practice and Leadership, University of Technology Sydney Sydney, Australia https://orcid.org/0000-0002-7528-9713

G. M. Hassan School of Computer Science and Software Engineering, University of Western Australia Perth, Australia https://orcid.org/0000-0002-6636-8807

S. Grundy School of Chemical Engineering, University of New South Wales Sydney, Australia https://orcid.org/0000-0002-5364-6011

R. Haque School of Science, Technology and Engineering, University of the Sunshine Coast Sippy Downs, Australia https://orcid.org/0000-0002-8641-4479

S. Nikolic Faculty of Engineering and Information Sciences, University of Wollongong Wollongong, Australia https://orcid.org/0000-0002-3305-9493

M. Belkina College, Western Sydney University Sydney, Australia

https://orcid.org/0000-0002-6636-8807

¹P. R. Neal peter.neal@unsw.edu.au S. Lyden School of Engineering, University of Tasmania Hobart, Australia https://orcid.org/0000-0002-5364-6011

C. Sandison Faculty of Engineering and Information Sciences, University of Wollongong Wollongong, Australia <u>https://orcid.org/0000-0002-5475-9640</u>

Conference Key Areas: Digital tools and AI in engineering education; Building the capacity and strengthening the educational competences of engineering educators **Keywords**: Generative AI, Academic Integrity, Learning Design, Assessment, Professional Development

ABSTRACT

ChatGPT and other Generative AI tools (GenAI) have generated much commotion and confusion within academic circles. Many academics still need to understand the risk such tools pose on current assessment practices and how students can use them for improved grades/outcomes. Unfortunately, most studies are very generic, do not go into much detail, and are outside of the engineering education context. In addition, GenAI provides many opportunities to construct novel, authentic and/or personalised learning experiences for students.

The workshop's facilitators have completed two comprehensive studies evaluating ChatGPT's impact on engineering education assessment (Nikolic et al. 2023; Nikolic et al. 2024). While strengths, weaknesses and opportunities were outlined, it is only the starting point for much-needed conversations. This workshop introduced participants to various GenAI tools in the context of a variety of engineering assessment types and engaged participants with a range of experience levels. We guided and engaged participants in exploring GenAI capabilities, in applying a framework for evaluating security and opportunities in assessment and through this demonstrated a model for professional development that they can adapt to their own institutions.

1 INTRODUCTION

The advent of user-friendly generative artificial intelligence (GenAI) tools is disrupting the conceptualisation and practice of engineering education. These tools offer users the opportunity to generate computer programs, reports, data analysis and even technical drawings with a few lines of natural language and limited knowledge of the field. Thus, educators, institutions and governments, accreditation and professional bodies are being forced to reconsider what it means to demonstrate learning, to assure competent graduates and even what learning is required.

There is significant current debate about how GenAI should be considered and employed. This debate joins a historical discourse that has paralleled the development of artificial intelligence (Bearman, Ryan, and Ajjawi 2022). Some frame these tools as a source of harm or disintegrity (Thorp 2023; Fischer 2023), while others see them as the dawn of a "post-learning era" (Siemens 2020). Then, of course, there is a significant spectrum of nuanced views in between (e.g., Lim et al. 2023; Yusuf, Pervin, and Román-González 2024). The tools obviously have an impact on the integrity of many current assessment practices (Nikolic et al. 2023) as well as providing exciting possibilities to create new learning opportunities (Chauncey and McKenna 2023). At the same time, concerns are being raised about these tools, for example about how they may impact student agency and metacognition (Abbas, Jam, and Khan 2024; Darvishi et al. 2024).

In our first benchmarking study (Nikolic et al. 2023), we evaluated the performance of ChatGPT 3.5 against assessment tasks drawn from a wide variety of engineering subjects. We discussed the implications for how different assessment types can be used to offset the risk of GenAI undermining assessment integrity and explored opportunities for using GenAI to support student learning. Given the rapid advances in GenAI's capabilities, we have updated and expanded our work to assess a wider range of GenAI and developed an Assessment Security & Opportunity Matrix for analysing different assessment types (Nikolic et al. 2024). In parallel, we have completed a systematic literature review which synthesises evidence from recent research literature evaluating teaching interventions using GenAI (Belkina et al. 2024). During the workshop we briefly shared key insights from these studies in the workshop and equipped participants to apply these insights to their own assessments and teaching practice.

2 RATIONALE

There is a massive need for professional development of engineering educators to meet the challenges and opportunities presented by GenAI. Typically, there is a long lead-time between the take-up of new technology by innovators and early adopters, and its translation into the practice of the vast bulk of educators. With clearly identified weaknesses in current practice, and ongoing threats from constantly improving GenAI tools, educational leaders and leading educators must ensure that all their colleagues are able to navigate GenAI-enhanced education.

This workshop sought to empower attendees with the capacity to assess the impact and opportunities of GenAI on their teaching practice, as well as demonstrating a transferable model for the professional development of teaching staff in their own institutions.

3 LEARNING OUTCOMES

The objective of this workshop was that participants would be able to

1. Perform a risk analysis of their current assessment practice with respect to GenAI using an Assessment Security & Opportunity Matrix,

- 2. Identify opportunities for the integration of GenAI in their educational practice, and
- 3. Disseminate professional development for their colleagues on GenAl in assessment practice.

4 WORKSHOP PLAN

The workshop was designed for participants with a wide range of experience with GenAI. It was organised into four parts:

- 1. Welcome and introduction (10min). After welcoming and organising them into groups, the participants responded to an online open-ended prompt: "What do you think of when you hear AI and Assessment?". The presenters explained how the workshop was designed as a model for professional development, and then gave a mini lecture explaining how GenAI works, how to interact with it, and good practice approaches to writing prompts and interpreting output. The mini lecture was designed to help less knowledgeable colleagues get up to speed.
- 2. Exploring GenAl capability to complete assessments (20min). Participants were introduced to the evaluation process employed in Nikolic et al. (2023) and then worked in pairs to evaluate one of three assessment tasks using ChatGPT. Participants were also provided with an example prompt for their task. This was followed by small-group discussion at tables and then a debrief in plenary.
- 3. **Managing GenAl risks in assessment (20min).** Following on from the **Part 2** discussion, participants were presented with the Assessment Security and Opportunity Matrix developed by Nikolic et al. (2024) and its associated risk assessment method. Table groups then worked to evaluate the risk of GenAl impacting the assurance of student learning, as well as discussing controls to mitigate the risk. The discussions of the various groups were then shared and compared with the whole group.
- 4. **Conclusion (5min).** The presenters then summarised the workshop discussions and their connections to the concepts of assurance of learning, assessment validity and resilient assessment design. Finally, participants were encouraged to use the <u>workshop materials</u> at their own institutions as are you, dear reader!

5 WORKSHOP RESULTS

The workshop was attended by approximately 25 participants. These participants were distributed across 6 table groups. When asked for their thoughts about GenAl and assessment, there were 40 responses from 17 participants. The responses reflected both positive and negative sentiments. The negative sentiments centred on concerns about cheating and academic integrity, while the positive responses included new efficiencies and opportunities in assessment. The discussion provided a moment to acknowledge the way GenAl is problematised and/or idealised, and to recognise that we need to engage critically with new tools.

Participants then engaged in the evaluation activity and each table was assigned one of three assessments:

 Position Paper – students critically analyse and respond to a newspaper article about a resource development project published in the last 12 months,

- Internship Reflection students reflect on a critical incident in an internship and use it to inform future professional development, and
- Design Quiz students complete five short answer questions about the engineering design process.

Even though the first two tasks incorporate elements recommended to promote academic integrity (e.g. currency, higher order thinking, reflection, reference to class discussion), the participants found that all tasks could be completed to a relatively passable standard by ChatGPT. The best responses were found for the Quiz, followed by the Position Paper. Here participants noted it would be hard to "distinguish spurious responses" as they GenAI at least superficially met the criteria.

The Internship Reflection task prompted a broader range of opinions. Some found that because no reflective framework was specified it meant ChatGPT guessed how to write the reflection and sometimes these guesses were unsatisfactory. Others achieved a better response by modifying the example prompt to include a competency framework. While still others questioned whether it was really cheating if the student provided all the reflection elements as bullet points and simply asked ChatGPT to put that information in the correct structure.

The second activity, applying the Assessment Security and Opportunity Matrix, demonstrated the benefits of the risk assessment approach. Given the familiarity of risk assessments in other fields, participants were able to rapidly understand the process and apply it to evaluating assessment tasks. The Position Paper and Quiz tasks were considered highly likely to be compromised by GenAI giving them Very High risk ratings, while the Reflection task was rated Moderate. Similarly, the participants could apply the control hierarchy to redesign assessment tasks. For example, invigilating students writing the Position Paper, substituting an interactive oral assessment for the Reflection, or redesigning the Quiz so that students critique AI response or instead submit their prompt for answering the questions.

Overall, the workshop produced lively and animated discussions with all participants engaged in the activities. Several participants found the risk approach was beneficial – rather than panicking or being paralysed by the threat of GenAI, this process provides teachers with a constructive way forward ("how to rethink my assessment"). Subsequent feedback has indicated that participants have recommended the approach to colleagues – achieving the dissemination of professional development outcome. Another participant commented that "The workshop was absolutely brilliant in a few respects: Firstly, [they] had a really clever design that quickly got participants engaged in highly relevant and thought-provoking tasks. Secondly, ... they were able to [rapidly] connect us with the emerging findings of this project, and the focused challenges that are facing engineering educators."

6 SIGNIFICANCE

The rate at which GenAl technologies are advancing means there's a risk of a new GenAl digital divide arising between educators and incoming students, threatening the integrity of assessments and that curriculum loses relevance for workplaces where GenAl is ubiquitous. Thus, it's imperative we usurp the typical long-term propagation pattern of technology innovation, where early adopters are using technologies far in advance of them becoming common practice. This workshop

contributes to that imperative i by developing participants' understanding of the challenges and opportunities of GenAI in engineering education and equipping them to share those insights with their colleagues.

7 **REFERENCES**

Abbas, M., F. A. Jam, and T. I. Khan. "Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students." *International Journal of Educational Technology in Higher Education* 21, no. 1 (2024): 10. https://doi.org/10.1186/s41239-024-00444-7.

Bearman, M., J. Ryan, and R. Ajjawi. "Discourses of artificial intelligence in higher education: a critical literature review." *Higher Education* 86, no. 2 (2022): 369-385. https://doi.org/10.1007/s10734-022-00937-2.

Belkina, M., S. Nikolic, S. Daniel, S. Grundy, R. Haque, G. M. Hassan, S. Lyden, P. Neal, and C. Sandison. "Implementing Generative AI (GenAI) in Higher Education: A Systematic Review of Case Studies using LCF, SAMR and TPACK Frameworks." *Computers & Education* (2024): Under review.

Chauncey, S. A., and H. P. McKenna. "A framework and exemplars for ethical and responsible use of AI Chatbot technology to support teaching and learning." *Computers and Education: Artificial Intelligence* 5 (2023): 100182. https://doi.org/10.1016/j.caeai.2023.100182.

Darvishi, A., H. Khosravi, S. Sadiq, D. Gašević, and G. Siemens. "Impact of AI assistance on student agency." *Computers & Education* 210 (2024): 104967. https://doi.org/10.1016/j.compedu.2023.104967.

Fischer, J. E. "Generative AI Considered Harmful." Paper presented at the *5th International Conference on Conversational User Interfaces*, Eindhoven, Netherlands, 19-21 July 2023. https://doi.org/10.1145/3571884.3603756.

Lim, W. M., A. Gunasekara, J. L. Pallant, J. I. Pallant, and E. Pechenkina. "Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators." *The International Journal of Management Education* 21, no. 2 (2023): 100790. https://doi.org/10.1016/j.ijme.2023.100790.

Nikolic, S., S. Daniel, R. Haque, M. Belkina, G. M. Hassan, S. Grundy, S. Lyden, P. Neal, and C. Sandison. "ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity." *European Journal of Engineering Education* 48, no. 4 (2023): 559-614. https://doi.org/10.1080/03043797.2023.2213169.

Nikolic, S., C. Sandison, R. Haque, S. Daniel, S. Grundy, M. Belkina, S. Lyden, G. M. Hassan, and P. Neal. "ChatGPT, Copilot, Gemini, SciSpace and Wolfram versus Higher Education Assessments: An Updated Multi-Institutional Study of the Academic Integrity Impacts of Generative Artificial Intelligence (GenAI) on Assessment, Teaching and Learning in Engineering." *Australasian Journal of Engineering Education* (2024): https://doi.org/10.1080/22054952.2024.2372154

Siemens, G. "The Post-Learning Era in Higher Education: Human + Machine." *EDUCAUSE Review* 55, no. 1 (2020): 60-61.

Thorp, H. H. "ChatGPT is fun, but not an author." *Science* 379, no. 6630 (2023): 313-313. https://doi.org/10.1126/science.adg7879.

Yusuf, A., N. Pervin, and M. Román-González. "Generative AI and the future of higher education: a threat to academic integrity or reformation? Evidence from multicultural perspectives." *International Journal of Educational Technology in Higher Education* 21, no. 1 (2024): 21. https://doi.org/10.1186/s41239-024-00453-6.