

Generative AI in assessment

Guidance for staff

*Nigel Francis – Cardiff University – francisn10@cardiff.ac.uk
David Smith – Sheffield Hallam University – d.p.smith@shu.ac.uk*

Generative AI in assessment

Guidance for staff

Nigel Francis – Cardiff University – francisn10@cardiff.ac.uk

David Smith – Sheffield Hallam University – d.p.smith@shu.ac.uk

Introduction

Generative AI (GenAI) exploded into the media in November 2022 when OpenAI launched its latest version of ChatGPT (GPT3.5). Since then there has been a near-daily launch of new tools and features, which shows little sign of slowing down, leaving the higher education sector with little choice but to embrace these new tools.

N.B. Fundamentally, from an assessment standpoint, there are no independently validated tools that can reliably and accurately detect GenAI-produced material.

This has huge implications when considering academic misconduct. We cannot reliably detect GenAI-generated outputs and have no way of proving, beyond any doubt, that a student has used an AI tool unethically unless the student admits to it. Established methods of looking at past submissions for changes in style and content will only hold short term. Students starting a course now will have full access to AI from the start, so we will have no point of reference.

Furthermore, GenAI is as incapable as it will ever be, and newer versions and paid-for versions offer vast improvements in capability. This has the potential to introduce a new digital divide between students who can afford to pay the monthly fee to subscribe versus those who cannot.

The misuse of generative AI raises serious ethical and pedagogical concerns. It undermines the integrity of the assessment process, devalues genuine learning and achievement, and can lead to a loss of trust in the educational system. We need to accept that students are aware of these tools, and some may choose to use them. Importantly, however, the use of AI is not academic misconduct unless used in such a way that it becomes unethical or is used to generate complete assignments that students then pass off as their own. Equally important is that some students may not wish to engage with AI tools, which normally require some personal data to be provided to sign up. Therefore, no assessment should mandate the use of GenAI tools to avoid disadvantaging students who do not wish to use them. Enterprise versions of these tools may circumvent this last concern.

What assessment modalities are at risk?

Currently, the most accessible GenAI tools are restricted to text-based outputs or image generation, however, this exposes a wide range of our current assessments to being AI 'ed. Any take-home assessment is susceptible in whole, or in part, to GenAI tools; this is particularly true of written assignments.

This means for most of our assessments that the final product becomes less important and the process by which that product was generated becomes far more important. If you are in doubt about the robustness of an assessment to being AI 'ed then you could enter the assessment into a GenAI tool and review the output.

Essays

Tools like ChatGPT will produce human-like text outputs that students could submit as their own without engaging in any critical thinking. Iterative prompting can dramatically increase the output quality, even mimicking the student's writing style. GenAI can be used to redraft existing text to bypass plagiarism checkers, with tools like Consensus and Jenni being capable of inserting genuine research citations.

Lab reports and data interpretation

GenAI is capable of fabricating data that fit defined research trends. Some AI tools allow CSV or PDF files to be uploaded, which opens the scope for automated data analysis and interpretation, bypassing the critical thinking and understanding typically required. Many tools, including ChatGPT, Claude and Bard can write or debug code in common programming languages like Python or R to visualise the analysed data. Data analytics can then be performed using natural language prompts to integrate the data.

Paper critiques

PDFs can be uploaded to some GenAI tools, which can be asked to compare and contrast the studies. Some tools allow for the interrogation of PDFs by asking guided questions or can rephrase the contents in a different manner, for example producing summaries for a lay audience. Additionally, this can include suggestions for future work albeit the suggestions may not be practical.

Literature reviews

Tools like Claude allow for multiple papers to be uploaded in various formats. The tools can then be asked to summarise the papers for common themes and ideas producing a synthesised review. Here, again, iterative prompting can dramatically enhance the quality of the output.

Research Proposals / Grant Proposals

As with paper critiques, GenAI can synthesise ideas from across multiple papers to suggest potential experimental designs. Suggestions are likely to be accurate but experimental approaches may not be optimal or will be impractical from a time or cost perspective.

Presentations

Students can use GenAI to write a script for a presentation, and tools like Gamma can create entire presentations from simple text inputs. If not being presented in person then it is hypothesised that GenAI could potentially be used to synthesise a student's speech in real-time during a remote oral examination, allowing someone else to take the exam on their behalf. Some tools can also make it look as if someone is looking directly at a camera while they are reading from a (GenAI) prepared script.

Podcasts

As above, although speculative and technically challenging at this point, GenAI could be used to fabricate a student's voice. However, GenAI can create a conversational script between two individuals.

Posters

GenAI can suggest layouts, provide an outline, write content, fabricate, analyse, and visually present data for posters.

What strategies can be taken to make assessment more robust?

- 1) **Process-orientated assessments.** Shifting from product-oriented to process-oriented assessments, where students are evaluated on their approach, process, drafts, reflections, and engagement with the material, rather than just the final product.
 - a. Ask students to produce a research trail.
 - b. What databases/websites have they searched?
 - c. What search terms or keywords were used?
 - d. How did they decide whether to include or exclude references?
 - e. Did they use AI? If so, which tool(s) and what prompt(s) were used?
 - f. How were AI outputs checked for bias and factual content?
 - g. Ask students to reflect on producing the final output and what they have learned.
- 2) **Oral examinations or presentations.** Including viva voces and presentations as part of the assessment process, allows academics to gauge students' understanding and originality directly. AI might have been used to help produce the output, but students will still need to fully understand the context to answer questions about both their conceptual understanding and the process.
 - a. Individual or team presentations
 - b. In-person poster presentations.
 - c. If AI was used in the creation, ask students to explain how.
- 3) **Designing Assessments around Novel Areas of Study.** Creating assessments requiring unique, higher-order thinking, and problem-solving makes it more challenging for AI tools to generate appropriate responses. For example, designing fictional scenarios e.g., a made-up ecosystem with fabricated species so that students are unable to enter direct questions into GenAI tools.

- 4) **Implementing Open-Book and Authentic Assessments.** Embracing open-book and authentic assessments that reflect real-world tasks and problems reduces the incentive to cheat and focuses on applying knowledge rather than rote memorisation. The authentic use of AI in the creation then becomes part of the assessment.
- 5) **Collaborative Assessments and Peer Review.** Encouraging collaborative assessments and peer review, where students work together and evaluate each other's work, fosters a sense of responsibility and community.
- 6) **In-person assessments.** In some cases, in-person, closed/open-book assessments might be appropriate. However, it is important to use these sparingly as they are not an authentic assessment. The QAA has deemed the long-term use of this approach as a step backwards and unsustainable.
 - a. Anatomy spot tests
 - b. In-class assignments
 - c. Presentations
 - d. Oral vivas
- 7) **Competency-based assessments.** A good example of this would be practical skills, which AI cannot replicate. Fieldwork observations, ethical debates, or science communication assignments are further examples of this type of assessment.
- 8) **Portfolio assessments.** Assessments are built up over time with draft submissions, regular formative feedback and reflections on the learning process. The emphasis on process and growth, integration of diverse skills and frequent opportunities to provide formative feedback makes it challenging to use GenAI to complete these types of assessment.
- 9) **Physical artefact creation.** Assessment types that require the physical creation of artefacts can be enhanced by AI but not replaced. Such assessments would embrace the fabrication of models, notebooks created over time (i.e. laboratory notebooks, sketch pads), or hand-created artefacts. Within STEM this would include chemical reaction annotations and scientific drawings, such as botanical notations.

Culture

One potential way to address GenAI is to create a culture of ethical and transparent use of these tools within the academic community. Fostering a culture that embraces the use of GenAI will be a complex and ongoing process that requires an integrated approach. It will involve the development of clear policies, and opportunities to educate staff and students in the ethical and appropriate use of AI. Critically, it will require buy-in from all stakeholders from senior management to students to ensure educational values and integrity are at the heart of assessments.

Long-term

Thinking long-term, there may be a shift towards programme-level, or synoptic assessments across modules that can become incorporated into curriculum delivery. Programme-level assessments look to evaluate the programme-level learning outcomes i.e. the proficiencies that we wish a graduate to possess. Not only can these be highly authentic assessments but by assessing across modules the student assessment burden can be reduced.

Further reading

- [Reconsidering assessment for the ChatGPT era: QAA advice on developing sustainable assessment strategies](#)
- [Artificial intelligence \(AI\) in tertiary education](#)
- [Russell Group principles on the use of generative AI tools in education](#)
- [YouTube – Digitally Enhanced Education Webinars](#)