

What to do about assessments if we can't out-design or out-run AI?

educational-innovation.sydney.edu.au/teaching@sydney/what-to-do-about-assessments-if-we-cant-out-design-or-out-run-ai

By

July 12, 2023



Generative AI is improving at a rapid pace and is increasingly integrated into software that we use daily. It is becoming increasingly clear that there isn't going to be a way to design take-home assessments that are 'AI-proof'. Assessments that are authentic, or reflective, or personal, or contextualised, or multimodal, or localised – these can all be completed by new generations of generative AI. Perhaps this isn't all bad news – human-AI collaborations are likely to be the norm in the near future, so we need to start thinking about how to embrace AI as part of assessment. But, we still need to assure that students have achieved learning outcomes – how do we do this, whilst embracing AI?

Together with folk across the University of Sydney, we put together a quick guide to help coordinators ensure that assessment changes have longevity, even as AI progresses. The guide also emphasises the opportunities for us to reflect on the human side of teaching and learning, help students become better prepared for AI-augmented careers, and encourage the ethical, accountable, and transparent use of AI. Here is a summary of the contents of the guide; you are welcome to download the full guide. Note that the guide has been produced for use within the University of Sydney, but it has broader applicability across different institutions.

[Download the full guide](#)

Diagnosing and addressing impact

What are the immediate actions you need to take for your assessments?

Discover the capabilities. Run your assessment prompt(s) through generative AI, preferably using the more powerful models available (such as [GPT-4](#) for text generation, Adobe Firefly for image generation, etc). Ensure you play around with the prompts to properly explore AI's potential capabilities (see Appendix 1 in the [full guide](#)).

Evaluate student motivations. Students are more likely to undertake assessment [with integrity if they feel](#):

- **Autonomy:** having real choice about topic and mode, and seeing how the assessment *meaningfully* connects with their life and career.
- **Competence:** being supported to build confidence and skills gradually.
- **Relatedness:** feeling connected to teachers and peers and that they matter.

Balance assurance of learning and use of AI. All units need to have assessments that assure student attainment of learning outcomes – this is likely best performed in live, supervised settings. It's also critical to ensure we help students use AI productively and responsibly, which can be done by redesigning other assessments to address appropriate learning outcomes. The table on the next page provides guidance for balancing these.

Reduce the perceived workload and pressure. Assessments with clear instructions and criteria, have meaningful and appropriate challenge, that provide sufficient time for completion, and which help students develop confidence in their abilities (e.g. through structured drafts and feedback) will lead to more positive academic integrity outcomes.

Decide and communicate. It's important to differentiate AI use for *learning*, and AI use for *assessment*. Use this guide to determine what level of AI usage is appropriate in context, and clearly communicate this to students – including situations where AI use is not appropriate. Examples of appropriate use and wording about how to acknowledge AI use are provided in the [full guide](#).

The [full guide](#) has specific advice for different assessment types including take-home assignments, small continuous assessments, groupwork assessments, and exams and tests.

Redesigning assessments

Some level of assessment redesign is required across almost every unit to both manage the risk of generative AI *and* provide students with opportunities to engage with it productively and responsibly. In a world where AI is inescapable, assessments should both assure learning in secure settings, *and* adapt to the reality of AI in other settings, as appropriate to each discipline.

A 'two-lane' approach

The two-lane approach below emphasises balance between assurance, and human-AI collaboration. **The reality in any one unit will likely be a situation where some assessments lie in lane 1 in order to assure attainment of all learning outcomes, but most other assessments lie in lane 2.** Fundamentally, we want to develop students who are well-rounded and can contribute and lead effectively in authentic, contemporary environments (which will include AI), and also be assured of their learning. Therefore in this context, **it is important to privilege lane 2 assessments with a higher weighting than lane 1 assessments.**

We do not foresee a viable middle ground between the two lanes. It needs to be assumed that any assessment outside lane 1 (i.e. that is un-secured) may (and likely will) involve the use of AI.

Lane 1 – Examples of assured ‘assessment of learning’

- In-class contemporaneous assessment e.g. skills-based assessments run during tutorials or workshops
- Viva voces or other interactive oral assessments
- Live simulation-based assessments
- Supervised on-campus exams and tests, used sparingly, designed to be authentic, and for assuring program rather than unit-level outcomes

Lane 2 – Examples of human-AI collaboration in ‘assessment as learning’

- Students use AI to suggest ideas, summarise resources, and generate outlines/structures for assessments. They provide the AI completions as an appendix to their submission.
- Students use AI-generated responses as part of their research and discovery process. They critically analyse the AI response against their other research. The AI completion and critique provided as part of the submission. Appendix 3 in the [full guide](#) provides suggestions on how to assess this.
- Students initiate the process of writing and use AI to help them iterate ideas, expression, opinions, analysis, etc. They document the process and reasoning behind their human-AI collaboration. The documented process demonstrates how the collaborative writing process has helped students think, find their voice, and learn. The documented process is graded and more heavily weighted than the artefact. Appendix 3 in the [full guide](#) provides suggestions on how to assess this.
- Students design prompts to have AI draft an authentic artefact (e.g. policy briefing, draft advice, pitch deck, marketing copy, impact analysis, etc) and improve upon it. They document the process and reasoning: initial prompt, improvements, sources, critiques. The documented process demonstrates learning, is graded, and is more heavily weighted than the artefact. More information. Appendix 3 in the [full guide](#) provides suggestions on how to assess this.

We do not foresee a viable middle ground between the two lanes. It needs to be assumed that any assessment outside lane 1 (i.e. that is un-secured) may (and likely will) involve the use of AI.

An example of assessments across both lanes

In this example, students need to apply marketing strategy concepts in real-world scenarios, demonstrate their communication skills, and evaluate the effectiveness of different marketing strategies.

The lane 2 assessment might involve students collaborating with AI such as Bing Chat (which is internet-connected) to perform market research and competitor analysis, and other AI such as Adobe Firefly for the visual elements of campaign design. Students document their interactions with the AI tools, including the AI's initial market research and analysis and their critique and fact-checking processes to evaluate the AI's outputs. Students also critique whether AI provided novel insights and whether it missed critical factors. This is then presented live in class. The grading of the assessment is more heavily weighted on the documented process of critical co-creation (see Appendix 2 and Appendix 3 of the [full guide](#)).

The corresponding lane 1 assessment might involve a live Q&A after the presentation, where students need to defend their research and analysis through targeted questions. This can be made to simulate real-world business meetings, and helps to assure that students have met their learning outcomes of applying marketing strategy concepts and evaluating effectiveness of marketing strategies. Another lane 1 assessment might involve giving students an unseen case study of a company that has recently launched a new product; in a live, supervised setting, they need to evaluate the effectiveness of the marketing strategy and propose areas of improvement.

Using AI as part of assessment

The [full guide](#) provides examples of uses of AI in assessment, as well as wording that could be presented to students to explain this. It also provides examples of how students might acknowledge the use of AI.

Some examples include:

- Generating ideas for assessment: You may use AI tools such as <ChatGPT, Bing Chat, and other generative AI> to <brainstorm ideas and approaches> for completing your assignment.
- Creating media for assessment: You may use AI tools such as <DALL-E, MidJourney, Stable Diffusion, Adobe Firefly, and other image generative AI> to generate <images> that you use as part of your submission.
- Providing feedback on work: You may use AI tools such as <ChatGPT, Bing Chat, and other text-to-text generative AI> to seek feedback on your written work.
- Searching literature: You may use AI tools such as <elicit.org, perplexity.ai, and researchrabbit.ai> to find and summarise research articles. You then need to incorporate the scholarship yourself into your submission.

Detecting students' AI use

Software that purports to detect the use of AI-generated text is prone to false positives and false negatives. Research has suggested these tools are not reliable and may be biased against non-native English writers. In addition, ChatGPT does not 'know' whether it generated a piece of text – even if it may produce a convincing response when asked.

Our advice is that **you should not submit students' work to AI detection software yourself** – this is potentially a breach of student privacy and intellectual property.

More concrete advice

Testing your own assignment with generative AI

It's important to test your assignment(s) against generative AI to gauge what kinds of outputs could be produced. When prompting generative AI, it's important to remember that better prompts will yield better results – don't stop at the first prompt/response and dismiss AI if its initial response is not impressive.

Appendix 1 of the [full guide](#) has a step-by-step guide on how to properly try out your assignment prompts on AI.

Suggestions for rubric criteria that target higher order thinking skills

A marking rubric is a tool that allows teachers, markers and students to form a shared understanding of the specific criteria and standards used to make academic judgements. A rubric directs students' work by providing descriptions of the standards at different levels of achievement. Appendix 2 of the [full guide](#) has some sample wording that you should adapt for your needs.

Note that the rubric suggestions attempt to privilege the more human elements of writing and composition as part of assessment design and grading. It is becoming trivially easy for AI to replicate these elements if prompted the right way. Therefore, changing your rubric should **not** be the only change you make to assessment in response to AI.

Suggestions for rubric criteria that privilege the process of human-AI collaboration in assessment rather than the product

Appendix 3 in the [full guide](#) provides rubric examples that can be adapted for assessments that involve human-AI collaboration. These rubric criteria are designed to help you assess the process of learning and evaluate whether students have appropriately developed and applied disciplinary skills and knowledge when they are working with AI.

Approaches to viva voces, live Q&A, and other interactive oral assessment

Interactive oral assessments can be an authentic, secure, and engaging way to assess attainment of learning outcomes. Optimally, they are conversational in nature, as opposed to a question and answer oral test. They allow you to probe deeper understanding – it is often very easy to spot a student who doesn't understand a concept by their oral responses. We have provided [guidance on these assessment tasks](#), and provide additional guidance below for interactive oral assessments in the context of AI.

Appendix 4 in the [full guide](#) has more guidance about these assessments.

Tell me more

Check out our other [resources on AI and education at Sydney](#), including what it might mean for students to [collaborate with AI in assessment](#)

[Back to top](#)