Responding to generative AI for assessments in semester 2, 2023

Generative AI is improving at a rapid pace and is increasingly integrated into software that we use daily. This document guides coordinators to 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.

Diagnosing and addressing impact

For all assessment types

- **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).
- **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.

For 'submitted work' assessments[1]

- Generally: adjust marking rubrics to privilege critical thinking, use and integration of sources, etc (see Appendix 2)
- If marked artefact is written (e.g. essay, reflection, report)[2]: consider ways of assuring attainment of learning outcomes in supervised on-campus settings such as through draft sessions in tutorials
- If marked artefact is multimodal (e.g. presentation, podcast, video): include live Q&A session after delivery to assure student knowledge and skill (see Appendix 4)

For 'in class, small, continuous' assessments

- Mark student work that is mainly composed in class time
- Conduct live debates, Q&As, panel discussions, interactive oral assessments etc to assure knowledge (see Appendix 4)
- Consider adding compulsory weekly work-in-progress submissions

[1] Suggestions such as using unit-specific content, using more recent information, using personal reflections, and using students' experiences and contexts are fairly easy for AI to mimic, so we have not included these here.
[2] Suggestions such as turning these into multimodal assessments (e.g. presentation, video, poster etc) have minimal impact on assurance of learning – however, they may be more authentic and motivating.

For 'groupwork' assessments

- Establish shared understanding of expectations around AI use and shared responsibility for the submission
- Follow the guidelines for other assessment types
- Consider formative peer review of contributions

For 'skills based' assessments

- If these are supervised: this is lower risk
- If these are not supervised (e.g. use of software): additionally, assure key learning outcomes in a supervised environment

For 'final exam' and 'in semester test' assessments

- If online[3] (e.g. online quiz, online case study): this is high risk and should be redesigned to vivas (see Appendix 4) or other in-class assessment such as presentations, or a more assured form of assessment used
- If in-person: this is lower risk, but try to maximise authenticity of the assessment such as allowing devices, using case studies, allowing certain resources, etc – and note that these are not 100% cheat-proof

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.

The two-lane approach[4] 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.**

	Lane 1 – Examples of assured[5] 'assessment of learning'	**Lane 2 – Examples of human-AI collaboration in 'assessment as learning'**
Short term approaches	• Supervised on-campus exams and tests • Viva voces or other interactive oral assessments, if already in use • Increased use of class time for quizzes, drafting and other small assessments, coupled with contemporaneous tutor or peer marking and feedback	• 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 generate (or are provided) an AI response and suggest critiques applying a critical framework (e.g. clarity, cohesion, thesis, focus, evidence, organisation, language, references). The AI completion is provided as part of the submission. Need to consider if this is an authentic activity. • 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 provides suggestions on how to assess this.

[3] Suggestions such as using image-based questions are not a sustainable solution as AI advances. Browser plugins that read questions off LMS and provide AI-suggested answers make online quizzes highly unassured assessments.
[4] 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.
[5] AI use is already difficult to reliably detect, regardless of whether AI-generated content is submitted verbatim or AI is used in the process of generating an output.

	Lane 1 – Examples of assured[5] 'assessment of learning'	Lane 2 – Examples of human-AI collaboration in 'assessment as learning'
Longer term approaches	• 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[6], and for assuring program rather than unit-level outcomes	• 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 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 provides suggestions on how to assess this.

An example of assessments across both lanes

In this example[7], 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).

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

Examples of uses and wording

The below suggested wording should be modified depending on the form(s) of AI use that you deem appropriate. You can choose to start with a generic statement, followed by more specific wording based on the use(s) of AI you have allowed. For example:

> You are permitted to use generative AI to help you <insert learning activity and benefit> - this use must be acknowledged. For example, <provide some specific examples – see table below>.
>
> Do not post confidential, private, personal, or otherwise sensitive information into these tools. If you use these tools, you must be aware of their limitations, biases, and propensity for fabrication. Your use of AI tools must adhere to the Student Charter 2020, including upholding honesty, ethics, professionalism, and academic integrity. <If in a professional degree: Developing responsible use of AI is a critical part of professional

[6] In certain contexts, supervised time-limited live exams/tests can be authentic e.g. music listening, patient assessment.
[7] AI transparency note: GPT-4 was used to generate the idea for this example.

behaviour expected when you graduate.> Ultimately, you are 100% responsible for your assessment submission.

Example of AI use	Suggested wording
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.
Suggesting a structure or outline	You may use AI tools such as <ChatGPT, Bing Chat, and other generative AI> to help you <draft/write> <an outline for your work>.
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.
Tidying written language	You may use AI tools such as <Grammarly, Notion AI, ChatGPT, etc> to directly suggest readability improvements to your text in terms of grammar and expression.
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.
Writing a draft for later improvement	You may use AI tools such as <ChatGPT, Bing Chat, and other generative AI> to generate a draft artefact. You then need to <insert steps and documentation required> in order to demonstrate that you <insert learning outcomes>.

Suggestions on how to acknowledge the use of AI

How students should acknowledge the use of AI will depend on how they have used it. For lighter uses of AI, a suggestion like this might suffice:

> Use of generative artificial intelligence must be appropriately acknowledged. You can do this by <inserting a note at the end of your submission> where you need to <describe the AI tool(s) that you used, what you used it to do, what prompt(s) you provided, and how the output of the artificial intelligence was used or adapted by you>. This additional description does not add to your word count.

An example[8] of such an acknowledgement that a student may provide might look like:

> I acknowledge the use of ChatGPT (https://chat.openai.com/) to refine the academic language of my own work. On <date> I submitted my entire essay (<link to original document here>) with the prompt to <"Improve the academic tone and accuracy of language, including grammatical structures, punctuation and vocabulary">. The output (<link here>) was then modified further to better represent my own tone and style of writing.

If their use of AI is heavier, a different approach may be needed. This may involve having students document their prompt(s), and the process of critiquing and improving the AI output. We have provided guidance on this elsewhere.

Because the output produced by generative AI is non-deterministic (i.e. another person cannot obtain the same information as you did), **'citing' generative AI as a source is not appropriate**.

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.

[8] Adapted from Monash University LearnHQ

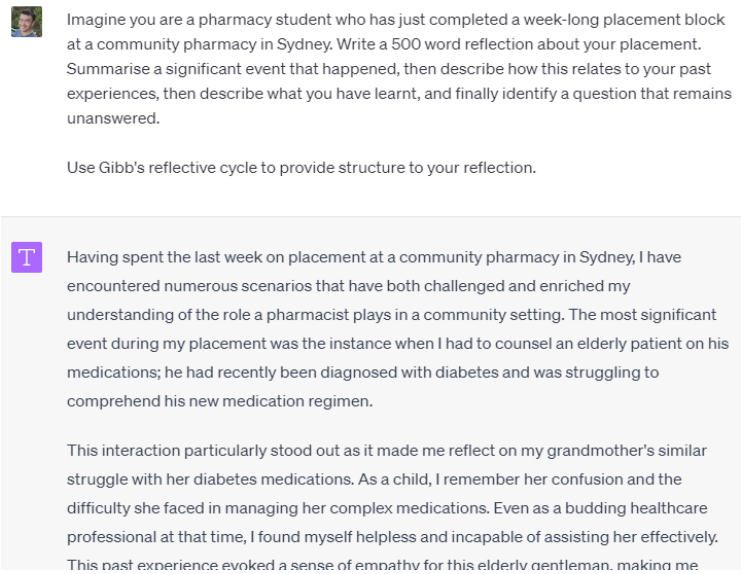
If you suspect that a student has used generative AI *inappropriately* for an assessment submission, you need to report this as a case to the Office of Educational Integrity for further investigation. **Do not submit the student's work to AI detection software yourself – this is a breach of student privacy and intellectual property.**

Appendix 1: 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.

Step 1 – Submit the assignment brief to an AI tool

1. Choose an AI tool (ChatGPT is recommended; if you or a colleague has access to the paid version of ChatGPT or another tool that provides access to the much more powerful GPT-4 engine[9], use that).
2. Paste relevant sections of the assignment brief into the AI tool. Ensure you include the task that students need to do, the requirements of the task, and background information that is provided to students. Ensure that no personal, private, or sensitive information is submitted to the AI tool.
3. Examine the response from the AI tool. Take note of the strengths and deficits of the AI response.

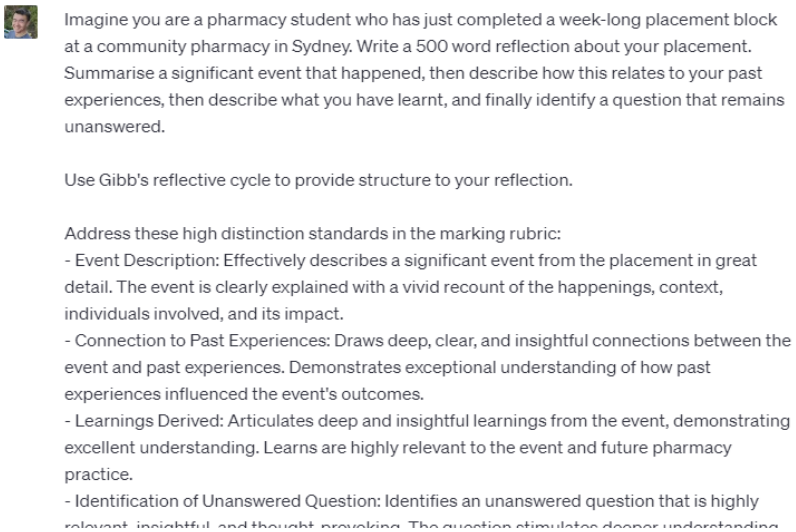


The screenshot to the right shows a simple personal reflective assignment prompt submitted to ChatGPT. Generative AI engines typically work better if they are given a persona; the example here starts the prompt with "Imagine[10] you are a pharmacy student who has just completed a week-long placement block at a community pharmacy in Sydney".

Step 2 – Improve the prompt to improve the response

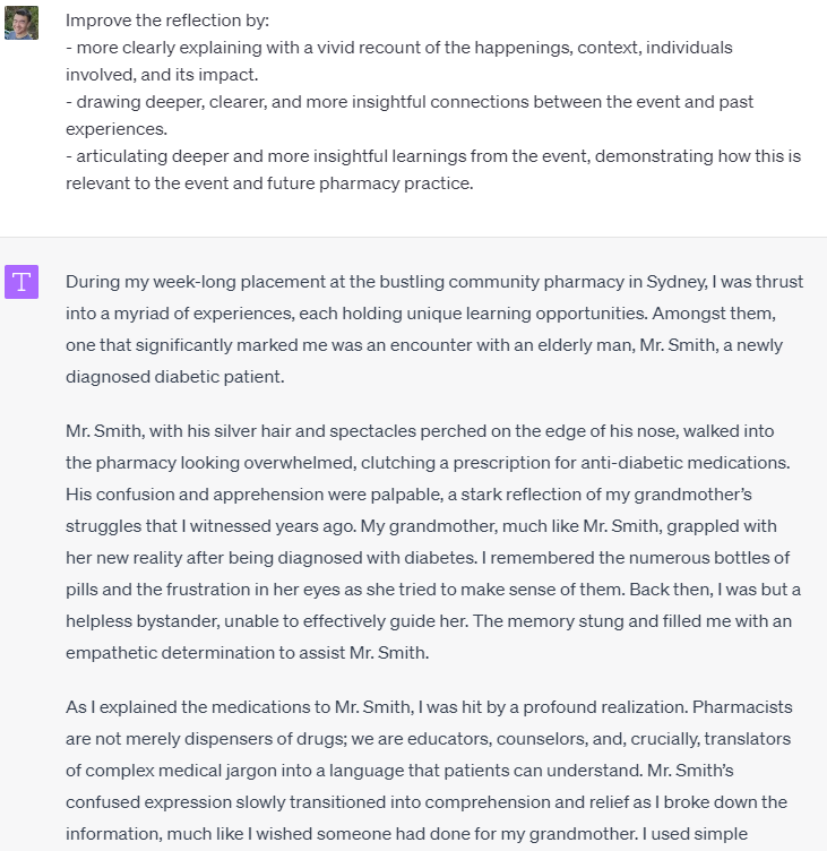
AI's response will improve if it is provided more information, and/or if further prompting is used to improve the initial response.

1. One approach is to adjust the initial prompt. For example, you could include the high distinction standards for each criterion in the marking rubric, demonstrated in the screenshot below. The HD standards are provided in a list to the AI as part of the prompt.



[9] Bing Chat from Microsoft, available at https://bing.com/chat when using the Microsoft Edge browser, also gives access to GPT-4 when set to 'creative mode'.
[10] The 'imagine' keyword also seems to encourage generative AIs to more fully engage in a prompt.

2. Another approach is to use follow-up prompts to improve the initial response. For example, you could provide a follow-up prompt that contains specific instructions for improvement, demonstrated in the screenshot below. In this example, the follow-up prompt asks the AI to address specific improvements targeted at the high distinction standards for three marking criteria.



Appendix 2: 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. We have provided some sample wording that you should adapt for your needs.

Note that the rubric suggestions below[11] attempt to privilege the more human elements of writing and composition as part of assessment design and grading. As we saw in Appendix 1, 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.

Application

At this level, students should be able to use knowledge in new situations. This involves applying theories to practical situations, using information to solve problems, or constructing models.

High Distinction	Distinction	Credit	Pass	Fail
Exceptional application of theoretical knowledge to novel, complex situations, generating original insights. Constructs comprehensive models or diagrams that effectively solve complex problems. Demonstrates a high level of adaptability and versatility.	Very good application of theoretical knowledge to new situations, showing ability to generate new understanding. Constructs clear models or diagrams that solve problems effectively. Demonstrates adaptability and versatility.	Good application of theoretical knowledge to familiar situations. Constructs basic models or diagrams that solve simple problems. Demonstrates some level of adaptability, though may struggle when situations become complex.	Acceptable ability to apply theoretical knowledge, but may struggle with unfamiliar situations. Constructs basic models or diagrams with limited problem-solving capability. Limited evidence of adaptability.	Fails to apply theoretical knowledge effectively. Does not construct useful models or diagrams. Struggles to adapt when faced with new or unfamiliar situations.

Analysis

At the analysis level, students should be able to break down complex concepts or problems into smaller parts to understand their structure. This involves recognizing patterns, identifying components, interpreting data, and distinguishing between relevant and irrelevant parts.

High Distinction	Distinction	Credit	Pass	Fail
Exceptionally discerns and articulates the structure and components of complex problems or concepts. Clearly identifies underlying patterns and relationships. Provides insightful interpretations of data, distinguishing	Demonstrates a very strong understanding of complex problems or concepts, effectively breaking them down into their components. Identifies key patterns and relationships. Interprets data accurately and distinguishes	Good deconstruction of problems or concepts into recognizable components. Identifies some patterns and relationships. Interprets data reasonably accurately and generally distinguishes	Acceptable ability to break down problems or concepts, identifying some components and patterns. Struggles with data interpretation and distinguishing between relevant and irrelevant parts.	Does not effectively break down problems or concepts. Unable to identify patterns or relationships and struggles with interpreting data. Fails to distinguish between relevant and irrelevant parts.

[11] AI transparency note: GPT-4 was used to draft the rubric criteria presented here.

High Distinction	Distinction	Credit	Pass	Fail
effectively between relevant and irrelevant parts.	between relevant and irrelevant parts.	between relevant and irrelevant parts.		

Evaluation

This involves making judgments based on criteria and standards. Students should be able to assess the value of theories, presentations, or outcomes. This includes critiquing ideas, selecting the most effective solution, or defending one's position.

High Distinction	Distinction	Credit	Pass	Fail
Exceptionally insightful and comprehensive critique based on clearly defined and appropriate criteria. Offers creative and highly effective solutions or defences. Demonstrates superior judgment in assessing the value of information or theories.	Provides a thorough critique, demonstrating a very good understanding of the appropriate criteria. Offers well-reasoned solutions or defences. Demonstrates good judgment in assessing the value of information or theories.	Provides a good critique using general criteria. Offers plausible solutions or defences. Demonstrates reasonable judgment, but may struggle when the situation is complex or nuanced.	Provides an acceptable critique with limited reference to criteria. Offers simple solutions or defences. Demonstrates basic judgment but may struggle with assessing the value of complex information or theories.	Does not provide a meaningful critique or offers a critique based on irrelevant or inappropriate criteria. Fails to offer viable solutions or defences. Demonstrates poor judgment in assessing the value of information or theories.

Creation

At this level, students are expected to put parts together to form a new whole, come up with new ideas, or create new things. This involves designing experiments, creating new models, proposing a new theory, etc.

High Distinction	Distinction	Credit	Pass	Fail
Demonstrates exceptional creativity and originality, producing innovative and ground-breaking ideas, models, or theories. Demonstrates a high level of independence, initiative, and original thought.	Demonstrates a very good level of creativity, producing new and practically applicable ideas, models, or theories. Demonstrates independence and initiative, taking ownership of the creative process.	Demonstrates a good level of creativity, producing new ideas, models, or theories within familiar contexts. Some evidence of independence and initiative in the creative process.	Demonstrates acceptable level of creativity, producing simple ideas, models, or theories. Lacks independence and initiative in the creative process.	Fails to demonstrate creativity. Does not produce new ideas, models, or theories. Shows no evidence of independence or initiative in the creative process.

Writing

Even though AI-generated writing can mimic human qualities, you may want to adapt this rubric criterion to emphasise to students that you are looking for deep, cohesive, and organised writing.

High Distinction	Distinction	Credit	Pass	Fail
The writing demonstrates a	The writing effectively	The writing effectively	The writing demonstrates some	The writing lacks understanding of

High Distinction	Distinction	Credit	Pass	Fail
deep understanding of the audience, purpose, and context, with appropriate style, voice, and language. The thesis is clearly stated and developed, and the writing is clear, cohesive, and organized, allowing readers to easily follow the flow of ideas.	communicates according to audience, purpose, and context, using appropriate style, voice, and language. The thesis is well-developed and well-organized, and the writing is clear and cohesive allowing the easy flow of ideas.	communicates according to audience, purpose, and context, using appropriate style, voice, and language. The thesis is well-developed and well-organized, and the writing is clear and cohesive, with occasional inconsistencies.	understanding of audience, purpose, and context, with mostly appropriate style, voice, and language. Inconsistencies in style, voice, language, and organization hinder the flow of ideas.	audience, purpose, and context, with inappropriate style, voice, and language. The thesis is unclear, weak, or absent, and lacks clarity, cohesiveness, and organization, making it difficult to follow the flow of ideas.

References

It is widely known that many generative AI models and tools can 'hallucinate' and make up references. Relying on the accuracy of references is **not** an adequate solution, as there are generative AI tools[12] that do draw from accurate scholarly sources. However, you may wish to adapt the following rubric criterion to encourage students to use sources well.

High Distinction	Distinction	Credit	Pass	Fail
All references are current, relevant, and specific to the discipline, demonstrating a deep understanding of the topic. At least two from the unit reading list are included, effectively supporting the arguments and assertions.	Most references are current, relevant, and specific to the discipline, demonstrating a solid understanding of the topic. At least two from the unit reading list are included, firmly supporting the arguments and assertions.	The references are somewhat current and relevant to the discipline, with some specific to the topic. At least two from the unit reading list are included, adequately supporting the arguments.	The references provided are limited in currency or relevance to the discipline. At least one from the unit reading list is included. They are used inconsistently or insufficiently to support the arguments and assertions, and may not be effectively integrated into the writing.	The provided references are outdated, irrelevant, and lack understanding of the discipline or topic. They are not included in the unit reading list or do not support the arguments.

[12] Such as elicit.org, the 'AI Research Assistant'

Appendix 3: Suggestions for rubric criteria that privilege the process of human-AI collaboration in assessment rather than the product

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. The rubric examples below can be adapted for assessments that involve human-AI collaboration. These rubric criteria[13] 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.

AI prompt design that demonstrates disciplinary expertise

How thoughtfully the student has designed the prompt(s) for AI and considered the complexity and clarity of prompts.

High Distinction	Distinction	Credit	Pass	Fail
Demonstrates exceptional and deep understanding of disciplinary concepts. The prompt design is sophisticated, unambiguous, addresses complex issues, and is optimally structured for AI, reflecting a deep understanding of the AI's limitations and capabilities.	Demonstrates very good understanding of disciplinary concepts. The prompt is clear, moderately complex, and largely effective for AI, showing good understanding of AI's limitations and capabilities.	Demonstrates a good understanding of disciplinary concepts. The prompt is understandable and effective but lacks consistent complexity or misses opportunities to leverage AI's capabilities or account for its limitations.	Demonstrates acceptable understanding of disciplinary concepts. The prompt may be unclear, oversimplified, or only partially effective for AI, showing insufficient understanding of AI's limitations and capabilities.	Fails to demonstrate understanding of disciplinary concepts. The prompt is unclear, overly simplified, or ineffective for AI, showing no understanding of AI's limitations or capabilities.

Critical evaluation of AI suggestions

How effectively the student evaluates and utilises AI suggestions, as in whether they simply adopt AI-generated content or make conscious choices about what to include.

High Distinction	Distinction	Credit	Pass	Fail
Critically and effectively evaluates AI suggestions, demonstrating exceptionally nuanced, evidence-based decisions about what to accept, modify, or reject, displaying a high-level understanding of how AI outputs compare to disciplinary knowledge.	Effectively evaluates AI suggestions with some evidence-based decisions about what to accept, modify, or reject, demonstrating a very good understanding of how AI outputs compare to disciplinary knowledge.	Evaluates AI suggestions with some critical thinking, but decisions are inconsistently supported by evidence, reflecting a good understanding of how AI outputs compare to disciplinary knowledge.	Evaluates AI suggestions, but critical thinking is minimal and decisions are often uncritical or without evidence, demonstrating an acceptable understanding of AI outputs in the light of disciplinary knowledge.	Does not effectively evaluate AI suggestions, with decisions appearing arbitrary or without justification, demonstrating no understanding of AI outputs and their link to disciplinary knowledge.

[13] AI transparency note: GPT-4 was used to draft the rubric criteria presented here.

Revision process

How the student has revised AI suggestions and demonstrated their critical thinking skills and disciplinary expertise.

High Distinction	Distinction	Credit	Pass	Fail
Provides insightful and critical reflection on where and why AI-generated content needed improvement. Demonstrates exceptional quality improvement and disciplinary expertise.	Offers a clear reflection on where and why AI-generated content needed improvement. Demonstrates very good quality improvement and good application of disciplinary knowledge.	Provides good reflection and shows some improvement in AI-generated content but lacks consistent demonstration of disciplinary expertise.	Provides acceptable reflection and some improvement of AI-generated content but only limited demonstration of disciplinary expertise.	Provides little to no reflection or improvement of AI-generated content and fails to demonstrate any disciplinary expertise.

Information and digital literacy

How the student has evaluated AI-generated content through relevant scholarly sources to enhance the rigour and reliability of the output.

High Distinction	Distinction	Credit	Pass	Fail
Exceptional evaluation of AI-generated content through the integration of high-quality, relevant scholarly sources. Critiques are thorough, enhancing the rigour and reliability of the output, and showing superior command over information literacy.	Very good evaluation of AI-generated content through the integration of relevant scholarly sources. Critiques are well-formed and add value to the reliability of the output, showing a very good command over information literacy.	Good evaluation of AI-generated content with some relevant sources but the critique lacks depth or consistency, demonstrating a good level of information literacy.	Acceptable evaluation of AI-generated content with minimal use of relevant sources and critiques are shallow, demonstrating an acceptable level of information literacy.	Fails to evaluate AI-generated content effectively or utilize relevant sources. Critiques are either absent or undeveloped, demonstrating a lack of information literacy.

Documentation and reflection on the co-creation process

How the student has recorded appropriate decisions and interactions with the AI co-pilot, and analysed the strengths, weaknesses, and future improvements to these interactions.

High Distinction	Distinction	Credit	Pass	Fail
Provides an exceptionally clear, ethical articulation of decisions and a deep insight into the role of AI in the co-creation process. Thoughtfully provides actionable	Very good articulation of decisions, which are clear and provide insight into the role of AI in co-creation. Offers some practical suggestions for future practice.	Good documentation of decisions and gives some insight into the role of AI in co-creation. Suggestions for future practice are sparse or generic.	Acceptable documentation of decisions and insights into the role of AI in co-creation. Lacks depth or forward-thinking.	Does not adequately document decisions or reflect on the role of AI in co-creation. Provides no meaningful suggestions for future practice.

High Distinction	Distinction	Credit	Pass	Fail
suggestions for future practice.				

Ethical considerations

Students' awareness of the reliability, biases, and other limitations of AI generated content.

High Distinction	Distinction	Credit	Pass	Fail
Demonstrates exceptional understanding of the reliability, biases, and other limitations of AI-generated content, and makes insightful suggestions for mitigating potential problems.	Shows very good understanding of the reliability, biases, and other limitations of AI-generated content, and suggests some ways to mitigate potential problems.	Good understanding of the reliability, biases, and other limitations of AI-generated content, but suggestions for mitigating potential problems are vague or incomplete.	Demonstrates acceptable understanding of the reliability, biases, or other limitations of AI-generated content. Suggestions for mitigating potential problems are generic or superficial.	Fails to demonstrate understanding of the reliability, biases, and other limitations of AI-generated content. Makes no suggestions for mitigating potential problems.

Appendix 4: Approaches to viva voces, live Q&A, and other interactive oral assessment

Interactive oral assessments[14] 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. Generally:

- The role of the assessment is for students to demonstrate what they know in a conversational style
- 10-15 minutes balances depth with strain for student and examiner
- Prepare students by informing them what to expect, and provide a sample (e.g. video of a mock oral assessment)
- Individual academic plans must be accommodated for
- The assessment should be run in a supportive and equitable way
- Use question styles that assess higher order skills such as application and analysis, instead of recall
- Design the interactive oral assessment experience to be authentic – modelling a realistic workplace/disciplinary scenario, and/or involving a realistic case or topic

Using interactive oral assessments to assure learning outcomes

- As a supplement to a 'lane 2' assessment: the interactive oral assessment follows a loosely-scripted set of questions that are related to the lane 2 assessment. The questions are written to unpack and fill knowledge gaps or misconceptions that may have been surfaced by the lane 2 assessment. The marking rubric would likely be the same as the lane 2 assessment, as the interactive oral assessment's role is to assure the same learning outcomes have been met.
- As a standalone 'lane 1' assessment: the interactive oral assessment follows a scripted set of questions that are related to key learning outcomes. The marking rubric is developed as usual, based on standards of achievement relating to the learning outcomes. During the conversation, the examiner uses follow-up questions to clarify and explore student understanding.

How to do this at scale

- You could consider a group-based interactive oral assessment, if the context suits. For example, students' individual written assessments could be marked ahead of the oral assessment, and key questions developed by the marker for oral follow-up. Students are then grouped and briefed on the topics they may be asked to expand upon in the oral. The oral itself runs as a facilitated group conversation, moderated by the examiner: the examiner may pose a question to one student, and another student may contribute a follow-up response or question.

[14] https://www.tandfonline.com/doi/full/10.1080/03075079.2019.1582015