

# Staff Guide



## Assessment and academic integrity in the age of artificial intelligence (AI)



Now available as a self-paced interactive guide on Amathuba



<https://bit.ly/3MtP0IV>

**CILT**  
University of Cape Town



**UNIVERSITY OF CAPE TOWN**  
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD  
CENTRE FOR HIGHER EDUCATION DEVELOPMENT



**Centre for Innovation**  
in Learning & Teaching



Updated August 2025 (originally compiled 2023)

© Copyright the [Centre for Innovation in Learning and Teaching](#)

Licensed under a [Creative Commons Attribution 4.0 International](#) licence

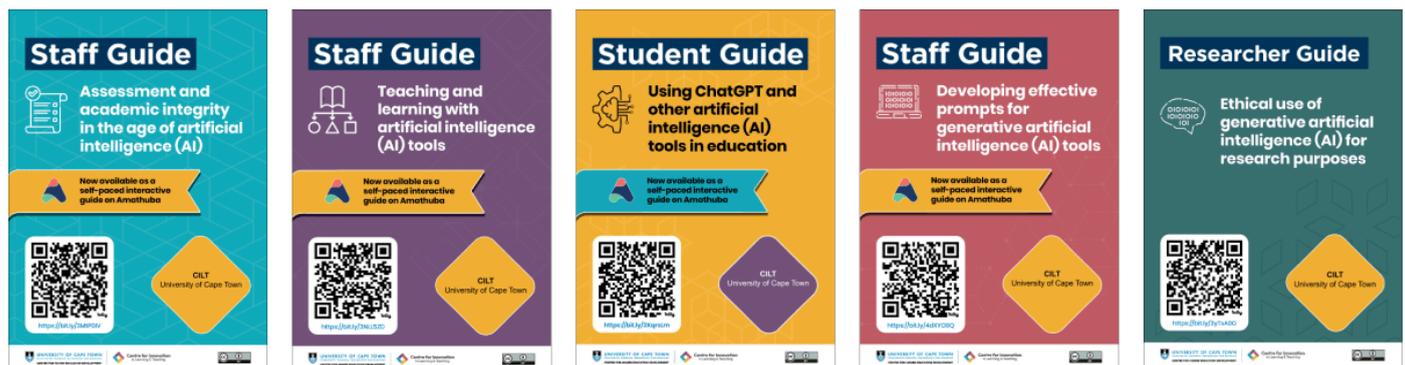
Published by the Centre for Innovation in Learning and Teaching, University of Cape Town



*This staff guide discusses issues impacting assessments following the wide availability of generative artificial intelligence tools such as OpenAI ChatGPT, Microsoft Copilot or Google Gemini. It considers approaches, practical suggestions and tools to mitigate threats to academic integrity. This is in support of the UCT Framework for [Artificial Intelligence in Education: Generative and Other AI in Teaching, Learning and Assessment](#) (June 2025).*

*AI is a fast-changing space, and the information in this guide is likely to become out-of-date quickly! This version was compiled in May 2023, updated in September 2023, March 2024, November 2024 and August 2025 and will continue to be updated regularly to include significant developments.*

View other [CILT guides in this series](#):





## Contents

<b>Introduction</b>	<b>3</b>
Communicating with your students	4
Academic integrity strategies	5
How students use generative AI tools for support	6
AI detection tools	6
Risks and concerns of using AI	7
Redesigning assessments linked to outcomes	8
Using prompts for preparation	9
AI-proof checklist for your assessments	10
How are AI detection tools' scores derived?	11
<b>References</b>	<b>13</b>

---

## Introduction

AI tools such as ChatGPT have sparked vigorous debates about assessments at universities. Concerns around plagiarism and cheating as well as questions when it is appropriate for students to use AI are raised. There have been calls for changes in the ways assessments are conceived, while others follow a return to paper-based only assessment strategies. How we respond will vary, so this guide presents a range of suggestions.

[ChatGPT](#), which stands for Chat Generative Pre-Trained Transformer uses neural transformer networks to generate text. It and other AI tools such as Gemini and Copilot are built on a category of AI tools known as 'Large Language Models' (LLM). These can perform various natural language processing tasks, such as, generate and classify text, provide answers to questions in a conversational style and translate texts from one language to another. These types of tools are commonly known as 'generative AI' which distinguishes them from other types of AI, already widely used in education (e.g., writing aids such as Grammarly or Quillbot). The use of generative AI, such as ChatGPT requires the user to insert a language prompt. ChatGPT then spits out the words and makes a prediction of the best answer to the prompt based on the information it was trained on.

AI tools like these might help to minimize some repetitive or administrative tasks. However, these tools can still pose risks of giving inaccurate information or producing a low-quality output. This despite the very confident way results are presented.

Students have responded to these easy-to-access, powerful capabilities by using the tools for assessments, raising concerns about academic integrity. While the AI landscape and its effects on higher education continues to be unpacked, there are attempts to circumvent the use of AI, such as by returning to more personal, handwritten and invigilated in-person exams.

There have also been calls for adapting assessments so that they discourage the use of AI. For example, if assessments involve a relatively simple fact recall that generative AI can respond to, there may be enhanced ways to assess which can reduce tools like ChatGPT's capacity to generate plausible answers. For instance, could the assessment also test students' abilities to compare, use and analyse this knowledge? Or could an assessment include an interactive short oral presentation that can be scalable and effective? These strategies may not always be appropriate or practical.



## Communicating with your students

---

AI literacies are becoming an essential skill. This involves developing an understanding of how the benefits must be balanced with the drawbacks. There is a need to equip our students with the necessary skills to navigate AI for their discipline.

At the beginning of your course, communicate to your students what you consider to be appropriate uses of generative AI tools for their assessments. There will be a spectrum across the university. This may include a requirement for a declaration that either such AI tools have not been used or that the use of these tools be [cited appropriately](#). If the use of AI is encouraged, you may want to specify tools or how/when students may use it.

The UCT Policy for the Prevention and Management of Academic Misconduct by Students (adopted in December 2023) defines cheating as:

... the practice of attempting to gain an unfair advantage. This includes accessing prohibited materials in an examination, making use of 'essay mills', language models such as chatbots (eg. ChatGPT and other large-language models or generative Artificial Intelligence), and any service or software that provides answers to assessments, or writes or re-writes assignments or parts thereof, other than software that detects and corrects spelling and grammatical errors. The only permissible instances of such practices are where they are explicitly permitted by the terms of the assessment instructions. In the case of group work, cheating includes students indicating that they have participated in group work when in fact they have not. ([UCT Policy for the Prevention and Management of Academic Misconduct by Students](#), p.3)

As per this policy, it is important to convey the following to your students:

- Specify whether or how generative AI tools may be used in a course or for an assessment. This may involve having a discussion with your students explaining the permitted uses of AI

and responding to any questions. Various examples of instructions have been compiled from other universities. Lance Eaton's '[Syllabi Policies for Generative AI](#)' and Tracy Moore's '[University Policies on Generative AI](#)' are collections of university policy and assessment instructions clarifying AI use.

- Specify how the AI tool should be referenced (e.g., APA style, [UCT Library guide](#)) and if further declarations should be made (i.e., including the prompts used).
- Include conditions in the course plagiarism declaration/honour pledge or in the assignment submission instructions (see below).
- Provide further guidance to students on using AI ([CILT student guide](#)).
- Explain the consequences of academic dishonesty and inappropriate use of AI.

The following adapted standard declaration can be used for your course with the additional clause on third party and software to generate assessments. You may use or alter this statement to suit the course's needs. It is not recommended to ban the outright use of artificial intelligence software as many existing tools (e.g., grammar and style tools) are AI-based. However, one can restrict the generation or creation of an assignment using AI software.

### **Student plagiarism declaration**

Example of adapting a plagiarism declaration to include using artificial intelligence software:

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. I have used the ..... convention for citation and referencing. Each contribution to, and quotation in, this essay/report/project/ ..... from the work(s) of other people has been attributed, and has been cited and referenced. Any section taken from an internet source has been referenced to that source.
3. This essay/report/project/..... is my own work, and is in my own words (except where I have attributed it to others).
- 4. I have not paid a third party to complete my work on my behalf. My use of artificial intelligence software has been limited to ..... (specify precisely how you used AI to assist with this assignment).**
5. I have not allowed and will not allow anyone to copy my work with the intention of passing it off as his or her own work.
6. I acknowledge that copying someone else's assignment or essay, or part of it, is wrong, and declare that this is my own work.

Many other forms of plagiarism declarations may be considered. The Liverpool University declaration specifies 'no commissioned production' that includes ChatGPT. Queens University, Belfast has a simple statement "I certify that the submission is my own work, all sources are correctly attributed, and the contribution of any AI technologies is fully acknowledged".



## Academic integrity strategies

Current generative AI can be used to ‘cheat’ on multiple choice quizzes, essays or coding assessments, and will undoubtedly continue to evolve. There are different ways to approach this challenge, as outlined by Michael Webb from JISC’s National Centre for AI in [A Generative AI Primer](#).

 Strategy	 Approach	 Challenges
<b>Avoid</b>	<b>Revert to in-person assessments where the use of AI isn't possible</b>	This moves away from authentic assessment and creates many logistical challenges
<b>Outrun</b>	<b>Devise an assessment that AI can't do</b>	AI is advancing rapidly and given the time between the assessment being set and it being taken, AI might well be able to do the assignment when it is taken
<b>Embrace and adapt</b>	<b>Embrace the use of AI, discuss the appropriate use of AI with students, and actively encourage its use to create authentic assessments</b>	Balancing authentic assessment and the use of generative AI with academic integrity is a challenge

Where possible, the ‘embrace and adapt’ approach seems more likely to be effective and successful going forward, although it still requires carefully considering the purpose and integrity of each assessment.



## How students use generative AI tools for support

Understanding more closely how students use generative AI tools for assessments is important. Popular uses of generative AI include summarising and paraphrasing long readings, generating ideas for assessment prompts, writing code, spelling and grammar checks (like Word and Grammarly), and generating practice questions for assessments. This may be analogous to how Wikipedia gives an introductory overview of a topic, which may be less of a concern or even encouraged.



## AI detection tools

---

There are concerns that students may be using AI tools to complete their assessments. In response, AI detection tools were developed claiming to flag when assignments could have been AI generated. Unfortunately, the reliability of these tools is generally poor. The reliability is certainly not comparable to a plagiarism checker using existing texts. The [UCT Framework for Artificial Intelligence in Education](#) (June 2025) clarified why Turnitin AI Scores will no longer be used:

**Responding to unacceptable AI use.** AI detection tools including the one available in Turnitin have been found to be unreliable in university contexts, often producing false positives or negatives or easily being subverted. Unlike traditional plagiarism checkers, AI detection tools rely on statistical methods to estimate the likelihood that a piece of text was generated by AI. This can lead to wrongful accusations of academic misconduct, damaging trust between lecturers and students, and creating an adversarial learning environment. At UCT, the Senate Teaching and Learning committee has decided to discontinue the use of the AI detection score in Turnitin effective 1 October 2025.

**For more discussion about AI detectors – see p.11**



## Risks and concerns of using AI

---

There are wider risks and concerns to be considered if you plan to use AI in your classroom that may not be immediately apparent.

**Inaccuracies in what is generated:** While AI outputs may come across as authoritative and convincing, its responses are based on next word predictions. The software has no real understanding. This leads to ‘hallucinations’ and allows it to produce disinformation (see [why ChatGPT produces inaccuracies](#)). It is important to evaluate this risk in the context of how you intend using generative AI.

**Bias in data fed:** AI follows the ‘garbage-in-garbage out’ principle whereby if the data it has been fed is biased, then its responses will reflect that bias. There are a number of biases that are present in AI data, such as, historical, representative, algorithmic, ranking, behavioural and social biases (read more about [bias and fairness in AI systems](#)). You need to be aware of the potential bias in AI outputs and critically evaluate the information it produces.

**Dependence/Overreliance:** Assignments serve the purpose of helping students learn and practice skills that we will need in society. While AI can help automate some tasks, it is important that students do not become overly reliant on it as this may hinder their ability to craft and think critically. Research suggests that being too reliant on AI can lead to loss of some important cognitive skills and prime us to think in a certain way (see [ChatGPT can homogenise our lives](#)).

**Ethical concerns:** Aside from the academic integrity concerns already discussed, there are a range of other ethical issues to navigate. [Data privacy](#), [environmental impacts](#) and [GenAI's potential to deepen inequality](#) are often not well understood. Exploring these questions with students can be a strategy for [developing critical AI literacies](#).

A typology of risks and concern organised from the most to least intentional is a way to consider developing responses:

- **Designated uses:** When AI is seen as a personal assistant, with understandings of the limitations and ethical concerns, few concerns are raised. Examples are given in the following sections.
- **Intentional misuse:** Where AI is intentionally misused includes cases of academic dishonesty, misinformation and deep fakes. This can involve possible academic misconduct which may not be easy to detect, as discussed above. Examples to adapt assessment to limit what AI might generate are discussed in the following section.
- **Accidents from use:** Cases where AI may hallucinate, hide biases, introduce inaccuracies, or raise ethical issues will become hard to avoid and identify. Some of these were discussed above.
- **Structural effects on society:** Some skills may no longer be valued while new skills are in increased demand. This could impact what gets assessed as discussed below.
- **Misaligned and power-seeking AI:** AI models will have bias and may exhibit goals not aligned with social justice.



## Redesigning assessments linked to outcomes

---

As outlined in the [UCT Framework for Artificial Intelligence in Education: Generative and Other AI in Teaching, Learning and Assessment](#) adopted in June 2025, promoting and ensuring assessment integrity is a key pillar of engagement with GenAI. The Framework recognises the many challenges being faced but stresses the importance of rethinking learning outcomes and the aligned redesign of assessment. Reviewing the original purpose of the course and considering the principles of assessment design (see the [Good Practice Guide for Assessment](#)) can provide a useful starting point.

While there are no simple solutions to the challenges of assessment redesign, there are established good practices which if engaged with can assist when considering the impact of AI on assessment (see [Do's, Don'ts and Don't Knows. Responding to AI in assessment in Universities](#)).

There are no quick fixes, so considering immediate and medium term changes is necessary. For an individual assessment task, Table 1 shows examples of opportunities to adapt assessments to target these higher forms of learning. Additional tips include reviewing the grading mechanism and rubrics and reducing the emphasis on recall.

Table 1: Assessment difficulty spectrum for AI (Adapted from Monash, 2023)

Difficulty level	Assessment type	Explanation of use
 <b>Easy</b>	<b>MCQ quizzes and questions involving recall</b>	Generative AI can easily provide output to factoid or low-level questions, especially on widely taught topics.
 <b>Easy</b>	<b>Generic short written assignments</b>	Generative AI can produce convincing essays and poems. For example, "Write a 1000-word essay on the history of Nelson Mandela".
 <b>Medium</b>	<b>Scaffolded submissions</b>	Creating a scaffolded assignment allows students to build on their previous work and feedback. Include pre-writing and drafting in the assignment process. Verifiable sources and citations should be required
 <b>Medium</b>	<b>Personalised or context-based assessment</b>	A simple essay can be made more challenging by introducing personalisation. Encourage personalisation where students are asked to draw on personal experiences. Ask students to write to a particular audience whose knowledge and values must be considered amplifying their student voice. Ask questions that would require them to give a response that draws from concepts that were done in class, in a lab, field trip or real-life experiences in their contexts.
 <b>Hard</b>	<b>Projects</b>	Projects involving real-world applications give students the opportunity for meaningful learning experiences.
 <b>Hard</b>	<b>Oral tests / exam with Q&amp;A / panels or discussions</b>	Synchronous oral assessment allows checking a student's understanding of their submitted work and gives the opportunity to interrogate their submission with follow up questions and discussion. While time intensive for larger classes, some have found ways to use these strategies.



## Using prompts for preparation

Generative AI has also been explored for routine assessment tasks and anticipating what students might find. There are, of course, limitations and concerns. Investigating these uses can help to understand how the AI tool works, what routine tasks can be automated, and how your students may use it.

### Testing existing assessments

Test your assignment briefs in an AI generative tool to see if it can be easily completed. Review the results and fine tune your prompts as a student would. Look closely at the output as while it may seem reasonable it may lack the finer details required from assignment briefs.

### Create quiz questions

Generative AI tools, given a lecture transcript or notes, can then be asked to generate multiple choice questions with answer options. It is usually good at generating the quiz format, the questions and options will require substantial editing and checking before being used in a course.

**Prompt:** "Read the following transcript from a lecture video and write 10 multiple choice quiz questions with 4 distractors each and feedback. Indicate the correct answer. This quiz should be pitched at a university level."

This is automated in the Amathuba Quiz tool when selecting the option to generate questions with AI. Doing it yourself may allow greater control. Suggestions for improving prompts include avoiding making them overly complicated as the output will become confused. One may also suggest to ChatGPT or similar tool to 'take the task step by step' and to 'rethink' earlier responses if the generated output is not satisfactory.

### Create rubrics

ChatGPT, Copilot or Gemini can be asked to generate a marking rubric. The prompt could include details such as the total marks and details of the task being assessed. This short video shows how to [Create a Marking Rubric With AI \(ChatGPT\)](#).

**Prompt:** "Generate a rubric out of 12 for a student assignment that requires them to create an app on Fliplet. These are university students studying in an education discipline."

### Generate writing prompts

ChatGPT or similar tools can generate possible writing prompts or sample data for a student assignment. Make sure to include the topic and theme in your prompt.

**Prompt:** "Generate five essay topics on using AI for teaching school mathematics. The essay should be 1000 words in length and use an example of how AI is used in schools"



## AI-proof checklist for your assessments

An AI misuse or proof checklist (Table 2) may assist when planning new assessments. No assignment can be made completely "AI-proof", but clear communications and reducing the likelihood that students could use AI tools inappropriately can be planned ahead. This is not to underestimate the time and potential resource implications of making changes to AI-proof assignments.

Table 2: AI proofing checklist (Adapted from Turnitin, 2023)

Criteria	
Does the assignment brief make explicit your institution and course's academic integrity policy, especially regarding the use of AI text tools?	
Does the assignment brief communicate the acceptable and unacceptable limits of using generative AI tools for the student response?	
Does the assignment brief require critical thinking or reasoning? Does the assignment	

encourage/require student voice?	
Does the assignment brief require the student to incorporate personal stories and/ or authentic situations?	
Does the assignment brief require a list of verifiable sources and/or citations? Are students asked to include a reflection or rationale for their approach to the assignment solutions?	
Have you instituted checkpoints to review outlines or drafts throughout the course, rather than focusing on a final submission only?	
Have you included time for peer reviews and/or discussions on learning activities throughout the course or assignment?	
Have you run the assignment brief through an AI generative tool?	

### How are AI detection tools' scores derived?

AI text detectors do not work in the same way as plagiarism checkers. Generative AI built on large language models is designed to produce text with the same statistical properties of natural language. This makes the detection problem a technical arms race as generative AI continually improves. Generative AI rarely outputs text that is verbatim copies of existing text, which is the basis of traditional plagiarism detection. A plagiarism checker such as Turnitin checks human generated text against other human generated text. Approaches to identify text generated by generative AI are thus done statistically. AI detectors attempt to estimate the *probability* of text being produced by AI. Some detectors are trained on datasets of human-written text and AI-written text to predict the probability that any given text is written by AI. Other detectors work by looking for features within the text, such as linguistic patterns, repeated words or structural patterns. Using multiple different AI detectors is unlikely to improve accuracy substantially.

A statistical approach to detecting AI generated text will likely continue to produce both false positives and false negatives and cannot decisively provide evidence of academic integrity violations. Currently such methods could easily lead to unfairly identifying academic integrity misconduct, creating mistrust between the lecturer and students, and damaging the learning environment.

The concerns with using plagiarism detection or originality tools to identify AI generated text are thus:

- There is a risk of falsely accusing students of plagiarism because detection errors are high.
- Some students will be able to evade the detection tools using other technologies.
- Assume any detector will be unreliable given the rapidly evolving changes in AI.
- Sharing student data by uploading it into a detector that is managed by a third-party company with unknown privacy and data usage policies is ethically and legally risky.

In an empirical review of AI text generation detectors, [Sadasivan et al. \(2023\)](#) found “that several AI-text detectors are not reliable in practical scenarios”. In a research study [Ibrahim et al. \(2023\)](#) state that “current AI-text classifiers cannot reliably detect ChatGPT’s use in school work”. The widespread use of other types of AI, such as Grammarly or Quillbot, further undermines accurate detection. The researchers found it was relatively easy to evade detection by running “the ChatGPT- generated text through Quillbot— a popular paraphrasing/ rewriting tool utilized by

students worldwide” resulting in rendering “these algorithms futile, failing to detect 95% of ChatGPT answers.”

There are freely available tools that given AI generated text will reduce the chances of this being detected as AI generated. These are known as AI content detection bypassers. These tools claim to be able to “humanise” your AI generated text, often through paraphrasing.

**These issues, there is an emerging consensus within the academic community that detectors are not reliable and may cause harm to particular groups of students.** AI detectors have been found to be more likely to [label text written by “non-native English” speakers as AI-written](#). In the study by Stanford researchers, further revealed a trend “where more literary language was classified as more “human”: enhancement of word choice in non-native English writing samples reduced misclassification, while simplifying native writing samples increased it, suggesting that GPT detectors are inadvertently penalizing individuals with limited linguistic proficiency.” (Liang et al., 2023).

These concerns informed the UCT decision to discontinue the use of the AI detection score in Turnitin. Several AI detection tools are available, including Turnitin, Copyleaks, Content at Scale and Originality.AI, however none of these has proved totally reliable. Initial claims from the vendors suggested high accuracy but real world testing and empirical is revealing these purported accuracy rates are unreliable and need further scrutiny:

- Turnitin AI detection has been discontinued at UCT.
- At various times Turnitin has backtracked on the initial claims [after wider usage and testing](#), and a number of universities have switched off the Turnitin detector including Wits and [Vanderbilt University](#). Interpreting the accuracy of Turnitin accuracy is still being investigated in university contexts, with evidence that Turnitin detection claims tend to be overstated.
- OpenAI has [shut down its own AI detector](#) due to low accuracy rates.
- A sample of the more commonly used detectors’ webpages or terms of reference indicate they should *not* be used as a primary mode to determine whether a text is AI-generated (source [AI detectors slide deck](#) from Torrey Trust).

# References

[UCT Framework for Artificial Intelligence in Education: Generative and Other AI in Teaching, Learning and Assessment](#) (June 2025)

Adendorff, H., Cilliers, F., Huang, C.W., Lester, S., Strydom, S., & Walji, S. (2026). Do's, Don'ts and Don't Knows. Responding to AI in assessment in Universities: A Practical Guide for Lecturers. Stellenbosch University & the University of Cape Town

Ibrahim, H., Liu, F., Asim, R., Battu, B., Benabderrahmane, S., Alhafni, B., et al. (2023). Perception, performance, and detectability of conversational artificial intelligence across 32 university courses. Scientific Reports. DOI: 10.1038/s41598-023-38964- 3

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. Patterns, 4(7), 100779. <https://doi.org/10.1016/j.patter.2023.100779>

Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the "originality" of students' work. Assessment & Evaluation in Higher Education, 1-14. <https://doi.org/10.1080/02602938.2024.2309963>

Monash. (2023). Generative AI and assessment. <https://www.monash.edu/learning-teaching/teachhq/Teaching-practices/artificial-intelligence/generative-ai-and-assessment>

Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). Can AI-Generated Text be Reliably Detected? arXiv preprint arXiv:2303.11156 <https://arxiv.org/abs/2303.11156>