



USO Project AI in Higher Education / Work package 4: AI literacy as a future skill

Responsible Chatbot use in higher education: implications for exam boards

A narrative literature review

31.01.2025

Dr. Eugène Loos

<https://www.uu.nl/staff/EFLoos>

e.f.loos@uu.nl

Utrecht University School of Governance

CONTENT

1. Introduction	3
2. Detecting chatbot generated texts: a mission impossible	12
3. Narrative literature review	18
3.1 Introduction	18
3.2 Method	20
3.3 Results	20
4. Conclusion and recommendations for exam boards	32
4.1 Conclusion	32
4.2 Recommendations	32
References	35
Appendix 1 Guidelines for detection of misuse of GenAI	39
Appendix 2 GenAI detectors	42

1. Introduction

We entered the chatbot era by the end of 2022 when ChatGPT became available for the broad public. Soon other chatbots followed (e.g., Google Gemini, Microsoft CoPilot Claude, and Perplexity; see also Fengchun & Holmes (2023/2024, p.10 for other examples), enabling citizens to create, use and access digital information 24/7, with large societal implications.

Chatbots such as ChatGPT also have an impact on university educational practices. February, 5 2023 Loos et al. (2023) asked ChatGPT to what extent ChatGPT can be used for education. ChatGPT answered as follows:

ChatGPT and other language models like it have the potential to be used in education in a variety of ways. Here are some examples:

1. Question Answering: ChatGPT can be used to answer questions in real-time, providing students with instant feedback and helping them to learn more effectively.
2. Tutoring: ChatGPT can be integrated into educational software to provide students with personalized, one-on-one tutoring sessions.
3. Content Generation: ChatGPT can be used to generate educational content, such as summaries, explanations, and study materials, based on input it has received.
4. Language Learning: ChatGPT can be used as an AI language tutor, helping students to practice speaking, writing, and comprehension in a foreign language.

However, it's important to note that while language models like ChatGPT have the potential to enhance education, they should not be used as a replacement for human teachers. AI systems like ChatGPT should be seen as tools to support and augment the educational process, not as a substitute for it.

Unesco (2023), p.10) presents a 4 step overview for the use of ChatGPT during the research process that is useful for research conducted by students: (1) research design, (2) data collection, (3) data analysis and (4) writing up. Utrecht University underlines that it is important that students should have ethical awareness while using ChatGPT:

The role of the university goes beyond teaching technical knowledge. Students must be equipped not only with the skills to handle generative AI, but also with ethical awareness. It is not enough that they understand how to use this technology; they must also consider its ethical and societal implications. Generative AI carries undeniable risks, from databias to privacy issues and the possibility of abuse. By being aware of these issues, the university can not only improve its own approach to generative AI, but also effectively convey this knowledge to its students. ([Problems with generative AI in education - Educational Development & Training - Utrecht University](#))

At the same webpage Utrecht University also lists the following ethical dilemma's: Sustainability, Data and trainer bias, Poor working conditions, Commercialization, accessibility and equality, Privacy, Intellectual property (see among others Kooli (2003) Loos and Radicke (2024) about ethical aspects), and adds two practical challenges: Hallucinations and Old data.

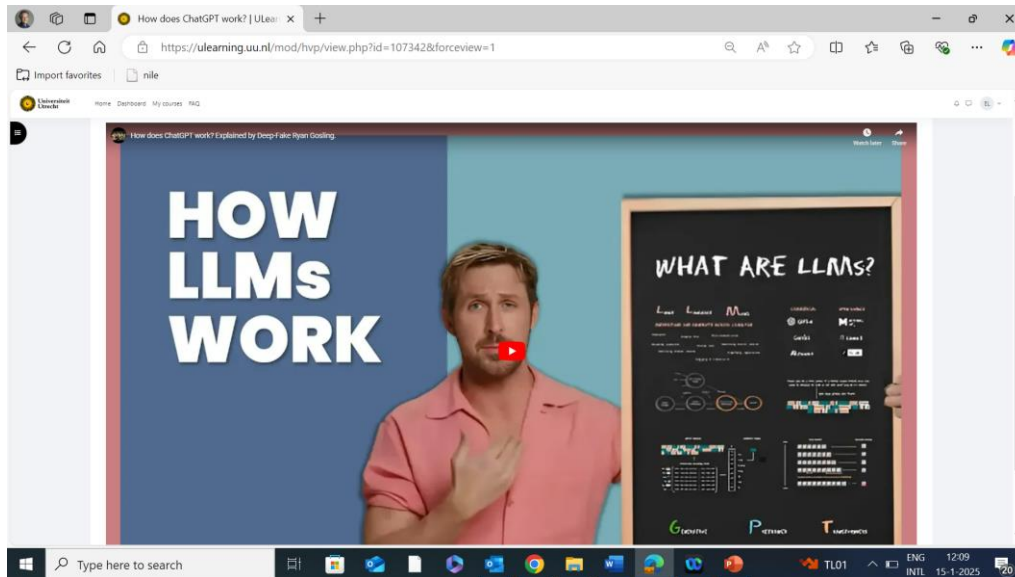
Unesco (2023, p.11) lists challenges and ethical implications: Academic integrity, Lack of regulation, Privacy concerns, Cognitive bias, Gender and diversity, Accessibility, Commercialization.

Fengchun, & Holmes (2023/2024, pp.14-17) present controversies around GenAI and their implications for education: Worsening digital poverty, Outpacing national regulation adaptation, Use of content without consent, Unexplainable models to generate outputs, AI-generated content polluting the Internet, Lack of understanding of the real world, Reducing the diversity of opinions and further marginalizing already marginalized voices, Generating deeper deepfakes.

Several studies conducted SWOT analyses related to the outcome of using chatbots (e.g. Farrokhnia et al., 2024; Mai et al., 2024), while Loos et al. (2023) asked ChatGPT to reflect on the Strengths, Weaknesses, Opportunities and Threats for educational use. Loos et al. (2023) also clearly showed that the underlying dynamics of conducting a “dialogue” with ChatGPT lead to Threats (e.g., so-called “hallucinations”) for students

using this tool and make suggestions how to tackle them. Hence, it is important more than ever to teach students how to enhance their critical thinking (Kuhn, 2019; Lai, 2011), by using what we call “future ready digital media literacy skills”. The question is not only how university students can find, assess and use reliable digital information (see also Wineburg & McGrew, 2016), but also how they can use ChatGPT and other chatbots (based on so called Large Language Models [LLM’s – see 💡 1) in a responsible way (Ferrari et al., 2023; Halaweh, 2023; Loos & Radicke, 2024; Williamson et al., 2024).

💡 1 How Large Language Models work



<https://ulearning.uu.nl/mod/hvp/view.php?id=107342&forceview=1>

Click on the link above, the log in by using your SolisID, then watch the explainer video

See also Fengchun & Holmes (2023/2024, pp. 8-13) about GenAI tools and how they work.

One could also ask how such tools affect the way we assess students' knowledge, skills and understanding as they can now easily use chatbot generated texts. Especially exam boards risk to be confronted with fraud cases when students use chatbots in a non responsible way, i.e. using a them without their lecturer's permission and not explaining how they proceeded with the use of this tool. In this regard it is important to take the stance of so called "tool criticism", for students as well as their lecturers in the institution where their educational practices take place:

(...) there are many easy-to-use tools for the collection, processing and analysis of data that require no knowledge of their limitations. Problematically, these tools are often assigned such values as reliability and transparency when in fact they are active mediators caught up in the epistemic process. (...) We underscore the urgency of this endeavour and its vital role for media and communication scholars.”
(van Es et al., 2021, p.46)

In this research report we aim to help exam boards to ensure that lecturers can continue to make assignments in a correct and fair way (see section 2), in other words how to enable lecturers to make assignments (see section 3.1 and 3.3) that can be used in the chatbot era? Formulating a strategy based on such guidelines to help lecturers to stimulate students' responsible chatbot use will play a central role to reach this goal.

Box 1. Utrecht University GenAI guidelines

Current guidelines for the use of generative AI at Utrecht University

[Guidelines for the use of generative AI - Corporate Identity - Utrecht University](#)

[Guidelines - Generative AI in Education - Intranet](#)

[Generative AI - Education - Utrecht University](#)

[Generative AI in education and the implications for assessment - Educational Development & Training - Utrecht University](#)

Guidelines for lecturers

[Generative AI - Education - Utrecht University](#)

[Responsible use of AI in your teaching - Intranet](#)

[Roadmap for lecturers - Generative AI in Education - Intranet](#)

Taskforce Responsible and Ethical Use of GenAI, Universiteit Utrecht, Faculty Humanities (see Appendix 1)

Guidelines for students

[Generative AI in Education - Students UU - Students UU](#)

<https://students.uu.nl/en/hum/practical-info/genai-during-your-study-the-guidelines>

Generative AI in education and the implications for assessment

[Generative AI in education and the implications for assessment - Educational Development & Training - Utrecht University](#)

Box 2. Dutch University GenAI guidelines

Erasmus University

[Generative AI Usage Guidelines | Erasmus University Rotterdam](#)

Maastricht University

<https://www.maastrichtuniversity.nl/file/guidelines-aipdf>

<https://www.maastrichtuniversity.nl/nl/file/beleidskader-generative-artificial-intelligence-12-12-20240pdf>

University of Groningen

<https://www.rug.nl/about-ug/organization/quality-assurance/education/artificial-intelligence-ai/?lang=en>

[Artificial Intelligence \(AI\) in education - EDU Support](#)

<https://www.rug.nl/about-ug/organization/quality-assurance/education/artificial-intelligence-ai/>

Radboud University

[Generative Artificial Intelligence | Radboud University](#)

[AI in education | Radboud University](#)

[Status of GenAI vision document | Radboud University](#)

Tilburg University

<https://www.tilburguniversity.edu/nl/intranet/portal-onderwijsondersteuning/ai-onderwijs>

<https://www.tilburguniversity.edu/nl/onderzoek/instituten-en-researchgroepen/chatgpt>

<https://libguides.uvt.nl/c.php?g=701954&p=5201565>

TU Delft

[AI chatbots in unsupervised assessment](#)

TU/e

https://assets.w3.tue.nl/w/fileadmin/Education_Guide/Content/Programs/Testing%20and%20assessment/AI%20Rules_TUe.pdf

University of Twente

[Guidelines for using AI during your studies at UT](#)

[Developments around ChatGPT](#)

Utrecht University

[Guidelines for the use of generative AI - Corporate Identity - Utrecht University](#)

[Generative AI - Education - Utrecht University](#)

[Generative AI in education and the implications for assessment - Educational Development & Training - Utrecht University](#)

UvA

[UvA policy on AI - University of Amsterdam](#)

[Beleidsmemo AI in het onderwijs - Universiteit van Amsterdam](#)

[GenAI & toetsing op vakniveau - UvA Teaching and Learning Centres \(TLC\)](#)

[AI tools and your studies - student.uva.nl](#)

VU Amsterdam

[Generative AI, Copilot and ChatGPT - Vrije Universiteit Amsterdam](#)

[How to deal with ChatGPT and Microsoft Copilot as a teacher - Vrije Universiteit Amsterdam](#)

WUR

[Generative Artificial Intelligence \(AI\) at WUR - WUR](#)

Box 3. GenAI policies and guidelines from international universities

Canadian universities and 35 from other countries: Policies and guidelines from Higher Education Strategy Associates on their AI Observatory

[AI Observatory | HESA](#)

Questions? Contact: info@higherstrategy.com

U.S. University Policies on Generative AI, Collection of university policies and websites.

[University Policies on Generative AI](#)

Questions? Contact: Tracy Mendolia: tmendolia@westernu.edu

Syllabi Policies for AI Generative Tools from international universities

https://docs.google.com/document/d/1RMVwzjc1o0Mi8Blw_-JUTcXv02b2WRH86vw7mi16W3U/edit?tab=t.0


Questions? Contact: Lance Eaton: lance.eaton@gmail.com

It is not the aim of this research report to discuss and evaluate the GenAI guidelines of these (inter)national Universities (for more information, see McDonald et al. (2025), Wang et al. (2024) for the USA and Jin et al. (2025) for a more global overview and Moorhouse et al. (2023) for an overview of the world's 50 top-ranking universities), but we would like to make a specific point that struck us. While Dutch Universities stress that students and/or lecturers should use GenAI in a responsible way, the stance they take toward the educational use of Gen AI differs. The following example illustrates this point. The VU Amsterdam states: “Can I use generative AI for courses at VU Amsterdam? The short answer to this question is that, as a student, you are not allowed to use generative AI **unless** the lecturer or examiner of your course indicates if and how this is allowed.” <https://vu.nl/en/student/examinations/generative-ai-your-use-our-expectations>, , while TU/e states: “TU/e aims to educate students within the context of their studies to become competent and responsible users of GenAI tools, aligning with academic practices,



attitudes, and core principles.”

https://assets.w3.tue.nl/w/fileadmin/Education_Guide/Content/Programs/Testing%20and%20assessment/AI%20Rules_TUe.pdf


In chapter 3 we will come back to the question of how to embed the use of chatbots in university educational practices. Though not a single university wants to forbid their use, it is important to decide in which cases its use should be limited to the minimum, allowed with restrictions, or propagated.

In general, the guidelines in  **Boxes 1, 2 and 3** are rather broad general GenAI policy frameworks on the one hand, and more specific guidelines for teachers and students on how to deal with GenAI in education on the other. In contrast, the question of what framework *exam boards* can use to encourage responsible use of *chatbots* by students is not really addressed. Therefore, this research report will focus on that point.

Chapter 2 first explains why it is impossible to detect chatbot use if it is not allowed for an assignment. In Chapter 3 follows a narrative literature review on responsible use of chatbots by students and teachers. And Chapter 4 presents the conclusion and recommendations.

We use  **Boxes** for additional background information and  for explainer video's.

2. Detecting chatbot generated texts: a mission impossible

One could ask what the problem is when students use chatbots for their assignments. The point is that such use may result “in a misrepresentation of their own performance in terms of knowledge, skills and understanding, which may result in the examiner no longer being able to assess the students' knowledge or ability correctly and fairly.” [Generative AI - Education - Utrecht University](#) (see  **Box 4**)

Box 4. Utrecht University's Education and examination regulations (OER)

The current Education and examination regulations (OER) already provides the right guideline which is also compliant with the introduction of GenAI:

“Fraud and plagiarism are defined as the actions or negligence of students that result in a misrepresentation of their own performance in terms of knowledge, skills and understanding, which may result in the examiner no longer being able to assess the students' knowledge or ability correctly and fairly.”

However, it may be possible, by way of example, to state that use of GenAI without stating so is considered fraud/plagiarism. Faculties can already do that independently: the examples mentioned of fraud and plagiarism are not compulsorily prescribed, programmes can add to them themselves. Moreover, the enumeration is not limitative or exhaustive. This follows from the provision "Fraud/plagiarism includes:" below:

Fraud includes: cheating during the test. The person who provides an opportunity to cheat is an accessory to fraud; sharing answers with others during a test; enlisting the help of third parties during a test; being in possession (i.e. having/carrying) during the test of aids (pre-programmed calculator, mobile phone, smartwatch, smartglasses, books, syllabi, notes, etc) whose consultation is not expressly permitted; having others or software create (parts of) a study assignment and offering it as if it were one's own work.

[Generative AI - Education - Utrecht University](#)

See also Anders (2023), Cotton et al. (2024) and Appendix 1 about cheating in the era of ChatGPT, and Kooli (2003) and Loos & Radicke (2024) about ethical issues in this domain.

Loos & Radicke (2024, p.2) state that “AI-assisted plagiarism [is] also called ‘AIgiarism’, according to Christian Nedelcu (<https://medium.com/@cristian.nedelcu/chatzero-is-the-best-tool-to-spot-aigiarmism-737846323985>, accessed 10.08.2023) a notion coined by the American venture capitalist Paul Graham. For more information about ‘AIgiarism’, see also Alex Hern in <https://www.theguardian.com/technology/2022/dec/31/ai-assisted-plagiarism-chatgpt-bot-says-it-has-an-answer-for-that> (accessed 10.08.2023) and Maretha Jonge in <https://www.leidenpedagogiekblog.nl/articles/chatbots-ben-je-als-docent-nog-wel-relevant> (accessed 10.08.2023).”

The next point to tackle is if it is possible to detect that a text is generated by a chatbot. Already December 11, 2022, shortly after ChatGPT’s introduction Sung Kim published the article ‘ChatGPT Output How to detect if the student used OpenAI’s ChatGPT to complete an assignment’ article about detection tools in *Geek Culture*. He states “ (...) Please note that these tools like everything in AI, have a high probability of detecting AI-generated text output, but not 100% as attributed by George E. P. Box “All models are wrong, but some are useful”. See also Khalil & Er (2023) in section 3.3.

Van den Bosch, professor Language, communication and computation at Utrecht University, also questions the validity of such detection tools and states that detectors are not reliable due to the occurrence of false positives and false negatives:


First of all, I would like to discuss the reliability of detectors of AI-generated texts, especially those used by both lecturer in evidence and students in defence. The first two were both published by OpenAI when GPT-2 was released in 2019 and GPT-3 in 2023, respectively. Other tools used include GPTZero, CopyLeaks and Writer AI Content Detector. As far as I can ascertain, these detectors (a) make around 1% to 5% classification errors [1] on validation tests performed by the creators themselves (but which are not verifiable), and (b) all of them provide an estimate of their own certainty in one way or another, which is not a sure indication about the correctness of their classification. Importantly, too, we do not know to what extent and how many false positives they generate (i.e., falsely classifying a human-written text as GPT-generated) or false negatives (i.e., classifying a GPT-generated text as human-written). We only get hints in the information provided, for example in the case of the detector for GPT-3: ‘We also bias the classifier to reduce false positives at the expense of more false negatives, because we believe false positives to be more costly than false negatives’. GPT-Zero gives a similar

indication, and at the same time allows the user to choose a different threshold in the model's certainty about the correct classification, which results in a different (but unknown) balance between false positives and negatives.

[1] Most detectors report on AUC, Area under the ROC Curve; see <https://www.sciencedirect.com/science/article/abs/pii/S016786550500303X> [Fawcett, 20025]. This metric is based on two components, false positive rate and true positive rate (recall); it is usually only reported on AUC and not the underlying components, even though it is relevant in this context.’

(...)

It is worth noting that in the future it may well become possible to still determine with high precision whether a text is AI-generated; when, for example, models such as ChatGPT / GPT-3.5 or GPT-4 with all model details are released afterwards, or when watermarks are secretly added to the text that are later released.” (internal document Utrecht University by Van den Bosch, 19.03.2024 – see Appendix 2 for more information)

For more information about the impossibility to make use of watermarks to detect chatbot generated texts in a reliable way, see  **Box 5**.

Box 5. Watermarks

“A watermark is a semi-transparent mark (a logo or text) that is embedded onto an image. The watermark signals who is the original author of the work. It’s largely seen in photographs and increasingly in videos.
Watermarking text in ChatGPT involves cryptography in the form of embedding a pattern of words, letters and punctuation in the form of a secret code. An influential computer scientist named Scott Aaronson was hired by OpenAI in June 2022 to work on AI Safety and Alignment. (...).
Something interesting that seems to not be well known yet is that Scott Aaronson noted that there is a way to defeat the watermarking. He didn’t say it’s *possible* to defeat the watermarking, he said that it *can* be defeated. **‘Now, this can all be defeated with enough effort. For example, if you used another AI to paraphrase GPT’s output—well okay, we’re not going to be able to detect that.’**
It seems like the watermarking can be defeated, at least in from November [2022] when the above statements were made. There is no indication that the watermarking is currently in use. But when it does come into use, it may be unknown if this loophole was closed.”
SEJ SEO (2023, January, 11). How The ChatGPT Watermark Works And Why It Could Be Defeated. *Search Engine Journal*. <https://www.searchenginejournal.com/chatgpt-watermark/475366/>

“Watermarking generative models consists of planting a statistical signal (watermark) in a model's output so that it can be later verified that the output was generated by the given model. A strong watermarking scheme satisfies the property that a computationally bounded attacker cannot erase the watermark without causing significant quality degradation. In this paper, we study the (im)possibility of strong watermarking schemes. We prove that, under well-specified and natural assumptions, strong watermarking is impossible to achieve.”

Zhang et al. (2023). Watermarks in the sand: Impossibility of strong watermarking for generative models. *arXiv preprint arXiv:2311.04378*
<https://doi.org/10.48550/arXiv.2311.04378>

“A small-scale study into watermarking suggests that this technique is feasible and show technical promise but should not be relied on as a solution to widespread use of artificial intelligence based tools by students. Alternative solutions are needed, including encouraging the educational community to work with artificial intelligence rather than against it. As such, the paper concludes by discussing seven potential areas for further exploration.”

Lancaster, T. (2023). Artificial intelligence, text generation tools and ChatGPT—does digital watermarking offer a solution?. *International Journal for Educational Integrity*, 19(1), 10.
<https://doi.org/10.1007/s40979-023-00131-6>

“Detecting GenAI-based plagiarism in written assignments: GenAI might allow students to pass off text that they did not write as their own work, a new type of ‘plagiarism’. GenAI providers are required to label their outputs with ‘generated by AI’ watermarks, while tools are being developed to identify material that has been produced by AI. However, there is little evidence that these measures or tools are effective. The immediate institutional strategy is to uphold academic integrity and reinforce accountability through rigorous detection by humans. The long-term strategy is for institutions and educators to rethink the design of written assignments so that they are not used to assess tasks that GenAI tools can do better than human learners. Instead, they should address what humans can do that GenAI and other AI tools cannot do, including applying human values such as compassion and creativity to complex real-world challenges.”

Fengchun & Holmes (2023/2024, p.28)

Does the fact that it is impossible to detect chatbots generated texts mean that from now on we are unable to assess the students' knowledge or ability? Certainly not! Chatbots may mean the death of the *traditional take home essay* (Marche 2022), but we agree with Rudolph et al. (2023, p.14) that in the chatbot era “Faculty can make use of these [AI] tools as a means to help students with writing and research, but not as a replacement for critical thinking and original work.”

A simple solution to the problem of students using ChatGPT would be to use physical closed-book exams where the students write by hand, using only pen and paper (Cassidy, 2023) – for online exams, proctoring/surveillance software can be used. However, such an approach to assessment (or at least an over-reliance on it) has been increasingly criticised as no longer contemporary, with students cramming less-than-useful information into their heads, only to forget much of it shortly after their examinations (Van Bergen & Lane, 2016). With a focus on graduate employability, the skill to ace closed-book exams seems rather irrelevant. Rudolph et al. (2023, p.13)

We also agree with Rudolph et al. (2023, p.13) that we should train students' skills to be prepared, after leaving university, for chatbot use. In a responsible way of course, to enable them to use "future ready digital media literacy skills" being aware of the Strengths, Weaknesses, Opportunities and Threats of this digital tool (e.g., Farrokhnia et al., 2024; Loos et al. 2023; Mai et al., 2024).

Liu & Bridgeman (2023, July 12) recommend to redesign assessments and explain:

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 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 [assured 'assessment of learning'] in order to assure attainment of all learning outcomes, but most other assessments lie in lane 2 [human-AI collaboration in 'assessment as learning']**. 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.**

See also Harvard's evolving collection of curated assignments that integrate AI tools from educators around the world ([Assignments – The AI Pedagogy Project](#)).


Another point is that it is important for lecturers to engage students in a dialogue about suspected academic misconduct ([How to Start a Conversation about Suspected Academic Misconduct - Student Accountability & Conflict Resolution](#)).


Please note that if the exam committee punishes a student who admits s/he used a chatbot in a non responsible way, but does not punish a student who denies this, this risks to lead to legal inequality.

In the next chapter we will use a narrative literature review to explain how exam boards can ensure that lecturers will be able continue to make assignments in the chatbot era. Formulating a strategy based on such guidelines to help lecturers to stimulate students' responsible chatbot use will play a central role to reach this goal.

3. Narrative literature review

3.1 Introduction

First of all we refer once again to  **Boxes 1, 2 and 3** in chapter 1 that can be used by exam boards to get insight into the GenAI guidelines from other (inter)national universities.

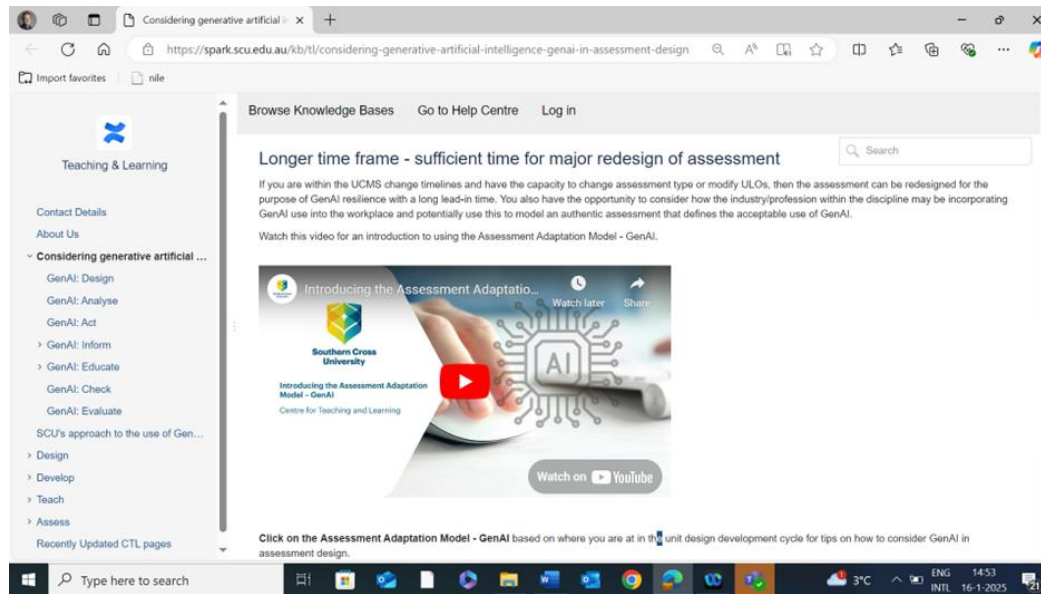
In this chapter we will therefore now present a narrative literature review (see  **Box 6**) to see if there are studies that can be used to enable exam boards together with their education directors to develop a strategy based on guidelines to help lecturers to stimulate students' responsible chatbot use . It is important to note that the use of chatbots in assignments is not per se forbidden, this depends on the learning goals.

We agree with Rudolph et al. (2023, p.13) who state:

Generally, we advise against a policing approach (that focuses on discovering academic misconduct, such as detecting the use of ChatGPT and other AI tools). We favour an approach that builds trusting relationships with our students in a student-centric pedagogy and assessments for and as learning rather than solely assessments of learning (Wiliam, 2011; Earl, 2012). The principle of constructive alignment asks us to ensure that learning objectives, learning and teaching and assessments are all constructively aligned. (Biggs & Tang, 2011)

For more information about how to put constructive alignment into practice we refer to [Generative AI in education and the implications for assessment - Educational Development & Training - Utrecht University](#) from Utrecht University. See also a You Tube The Assessment Adaptation Model - GenAI (AAM-GenAI) video from the Southern Cross University ([Generative Artificial Intelligence in Learning and Teaching - Southern Cross University](#)) to which this UU webpage refers:

💡 2. Toward redesigning assessments by adopting constructive alignment



The screenshot shows a web browser window with the URL <https://spark.scu.edu.au/kb/rt/considering-generative-artificial-intelligence-genai-in-assessment-design>. The page is part of a knowledge base for Southern Cross University (SCU). The main heading is "Longer time frame - sufficient time for major redesign of assessment". Below the heading, there is a paragraph of text explaining that if you are within the UCMS change timelines and have the capacity to change assessment type or modify ULOs, the assessment can be redesigned for the purpose of GenAI resilience with a long lead-in time. It also mentions the opportunity to consider how the industry/profession within the discipline may be incorporating GenAI use into the workplace and potentially use this to model an authentic assessment that defines the acceptable use of GenAI. Below the text, there is a video player thumbnail titled "Introducing the Assessment Adaptation Model - GenAI" from the SCU Centre for Teaching and Learning. The thumbnail shows a hand holding a pen over a document with a large "AI" logo. A "Watch on YouTube" button is visible below the thumbnail. The left sidebar contains a navigation menu with categories like "Teaching & Learning", "Contact Details", "About Us", and "Considering generative artificial ...". The Windows taskbar is visible at the bottom, showing the date and time as 16-1-2025, 14:53.

[Considering generative artificial intelligence \(GenAI\) in assessment design - Teaching & Learning](https://spark.scu.edu.au/kb/rt/considering-generative-artificial-intelligence-genai-in-assessment-design)

Note: Click on the *link* above to watch the explainer video.

Box 6. A narrative literature review

As chatbots are a relatively new tool, this research report is based on an exploration of papers from only the last two years. For this reason a narrative literature review is at the core of this research report. See: <https://towson.libguides.com/expert-reviews/narrative-literature-reviews#s-lg-box-26399333>
For more information about narrative literature reviews.

3.2 Method

We collected papers by using Google alert and ResearchGate, using the term ChatGPT from the moment on (by the end of 2022) this new tool got available for the broad public till the end of January 2025. As mentioned before our aim was to get insight into ways to help exam boards to ensure that lecturers can continue to make assignments in a correct and fair way, in other words how can they enable them to make assignments that can be used in this chatbot era. We collected papers addressing in the field of chatbot guidelines for higher education, allowing us to draw conclusions from the insights they offer for exam boards. We also used reports by Unesco and the Australian Government.

3.3 Results

The narrative literature below will show that *none of the papers* explicitly addressed the point of responsible chatbot use for university students *from the perspective of exam boards*, but the five Tables below will present papers addressing this topic *for lecturers and students*, allowing us to draw conclusions from the insights they offer for exam boards. Tables 1 – 5 address respectively the following points to tackle by exam boards: Chatting and cheating in the chatbot era, Development of chatbot guidelines, Constructive alignment, Training for students and lecturers, and Monitoring and validating GenAI systems for education. The studies we present in these five Tables are useful but we will also show that in most cases the results are not based on empirical studies but on other sources such as reports by Unesco and the Australian Government.

Table 1. *Chatting and cheating* in the chatbot era*

*see Cotton et al. (2024) for this creative title

Study/ Report	Country where the higher education institution is based	Explicitly mentioning implications for exam boards?	Based on empirical data and/or other sources	Main result
Anders (2023)	Armenia	No	Education/GenAI studies	"Students especially need to gain AI literacy in order to be competitive and effective in the job market, which is using AI more than ever. Yet all instructors must also develop strong AI literacy to be more relevant and effective in properly teaching these skills to students as a soft/power skill while also teaching the main subject of their instruction." (pp.1-2)
Cotton et al. (2023)	UK	No	Education/GenAI studies	"[The paper] suggests strategies that universities can adopt to ensure ethical and responsible use of these tools. These strategies include developing policies and procedures, providing training and support, and using various methods to detect and prevent cheating." (p.1)
Farrokhnia et al. (2023)	Netherlands	No	Education/ GenAI studies, ChatGPT, SWOT analysis	"(...) it comes as no surprise that Mike Sharples warns that 'GPT democratises plagiarism' (Welle, 2023). Students may utilise ChatGPT due to its promising capabilities without realising that it may lead to plagiarism. In addition, there is a high risk of plagiarism becoming more prevalent in academia." (p.8)
Khalaf (2024)	Oman (also study participants' country)	Yes	Attitudes Toward Aigiarism Questionnaire and Attitudes Toward Plagiarism Questionnaire	"Frequencies and percentages showed that 27% and 57% of students had positive attitudes towards plagiarism and aigiarism, respectively. No signifcant gender differences in aigiarism were detected. Attitudes towards aigiarism did not differ according to academic major (human-scientific) or lower and higher GPA"

Khalil & Er (2023)	Norway and Turkey	Yes	Education/ GenAI studies + “two popular plagiarism detection tools were used to evaluate the originality of 50 essays generated by ChatGPT on various topics.” (p.1)	“Our results manifest that ChatGPT has a great potential to generate sophisticated text outputs without being well caught by the plagiarism check software.” (p.1)
King & ChatGPT (2023)	USA	No	ChatGPT	College professors can design assignments to minimize potential cheating via ChatGPT by incorporating a variety of assessment methods that go beyond traditional essay writing.” (p.2)
Kooli (2023)	Canada	No	Education/GenAI studies	“Digital assessments are going to disappear and assessment methods need to be more creative and innovative. (...)The presence of AI systems and chatbots in education needs to be considered as an opportunity for development rather than a threat.”
Loos et al. (2023)	The Netherlands and Germany	No	Education/GenAI studies, ChatGPT	“(…) the possibility of plagiarism (also called Aigiarism (...)) is one of the biggest concerns regarding the adaption of ChatGPT in education. This has evoked responses varying from forbidding the use of ChatGPT or adapting the assessment of students’ learning, to creative ways to embedding it in educational practices (...). (p.8)
Loos & Radicke (2024)	The Netherlands and Germany, Platon	No	Education/GenAI studies, ChatGPT, Platon	“A first negative aspect of ChatGPT-3 is that it facilitates cheating, especially in educational and academic contexts, and rule-breaking. Obviously, cheating was already relatively common in education and academics (e.g., plagiarism). But engaging others to write one’s articles is more expensive than using ChatGPT-3 as a writing tool.”
Rasul et al. (2023)	Australia, Brazil, and India	No	Education/GenAI studies and experienced academics and practitioners	“Using ChatGPT as a tool for exploration and inquiry, students can actively construct their own knowledge and meaning, reducing the likelihood of academic dishonesty, such as plagiarism or cheating (Keith, 2022).”


Rudolph et al. (2023)	Singapore	No	Education/GenAI studies	"(...) specifically, develop policies and clear, easy-to understand guidelines for the use of language models in learning and teaching – the guidelines should include information on the proper use of these tools and the consequences for cheating (...)." (p.15)
Sullivan et al. (2023)	Australia, and news articles specifically from Australia, New Zealand, the United States and the United Kingdom	Yes	Content analysis to examine news articles (N=100) about how ChatGPT is disrupting higher education, concentrating specifically on Australia, New Zealand, the United States, and the United Kingdom	"Students reading multiple articles about students using ChatGPT to cheat may make them more likely to engage in that behaviour themselves."

All studies acknowledge that chatbot use in higher education leads to cheating and plagiarism. Most studies consider this as a danger for educational practices, but there are exceptions such as Kooli (2023) who states that: "The presence of AI systems and chatbots in education needs to be considered as an opportunity for development rather than a Threat." Many studies argue that it is important to embed chatbots in educational practices. It struck that most papers are not based on empirical research, most of them lean on insights from education/GenAI studies.

Table 2. *Development of chatbot guidelines*

Study/ Report	Country where the higher education institution is based	Explicitly mentioning implications for exam boards?	Based on empirical data and/or other sources	Main result
Unesco (2013)	-	No	Education/GenAI studies	<p>“Both the use of general GenAI platforms and the design of specific educational GenAI tools should be designed to enhance teachers’ understanding of their subject as well as their knowledge on teaching methodologies, including through teacher-AI co-designing of lesson plans, course packages, or entire curricula. The GenAI-assisted conversational teachers’ assistants or ‘generative twins of teaching assistants’⁵³ that are pre-trained based on data from experienced teachers and libraries, have been tested in some educational institutions and may hold unknown potential as well as uncharted ethical risk. The practical application processes and further iterations of these models still need to be carefully audited through the framework recommended in this Guidance and safeguarded by human supervision as exemplified in Table 4 [p.31].” (p.30)</p>
Lodge et al. (2023)	Australia	No	Experts from Education/GenAI studies	<p>“The most immediate concern is that students may use generative AI in assessment tasks, calling into question their personal learning attainment. It is challenging to design non-invigilated assessment tasks that preclude substantial use of generative AI, and it appears almost impossible to detect if these technologies have been used in the production of assessment products in a reliable way. Moreover, there is every possibility that what generative AI can produce will be of a passable quality for many assessment tasks, particularly as student capabilities for using these tools evolve over time. Therefore, it is necessary to address the nature of assessment in relation to generative AI directly to provide students and teachers with ways to approach this issue productively.” (p.1)</p>

<p>Fengchun & Holmes (2023/2024)</p>	<p>-</p>	<p>No</p>	<p>Education/GenAI studies</p>	<p>Use ChatGPT with care and creativity While some states and HEIs have blocked ChatGPT, most governments and HEIs are seeking ways to adjust to a world in which AI has become more widespread, accessible, and easy to use. In that sense, ChatGPT can be used but requires both care and creativity to ensure it is handled ethically and appropriately. Suggested ways forward include (...)</p> <ul style="list-style-type: none"> • Create opportunities for faculty, staff, students, and other stakeholders to discuss the impact of ChatGPT on the HEI and co-construct strategies to adapt and adopt to AI. (...) this type of engagement. • Introduce clear guidance for students and instructors about how and when ChatGPT can be used (and when it cannot). Such guidance should be negotiated with students and teachers, not imposed on them. • Connect the use of ChatGPT to course learning outcomes. This helps students understand how ChatGPT can support their learning and what expectations there are for them. • Review all forms of assessment and evaluation to ensure that each element is fit for purpose. This review may lead to replacing exams or other assessments with in-person assessments or altering the types of questions or exam formats that are used. • Review and update policies relating to academic integrity/honesty in relation to ChatGPT and other AI tools (...). • Train teachers, researchers, and students to improve the queries they pose to ChatGPT. As researchers have noted, ChatGPT is most useful when the inputs provided to it are carefully created • Train teachers, researchers, and students to improve the queries they pose to ChatGPT. As researchers have noted, ChatGPT is most useful when the inputs provided to it are carefully created (...)."
--	----------	-----------	--------------------------------	---

The three reports above could be used by exam boards to find inspiration to collaborate with their education directors to develop chatbot guidelines in their institution. They could also use the insights from  **Boxes 1, 2 and 3** in Chapter 1 for an overview of GenAI policies form

Dutch and international universities, and those from the studies listed in Table 1 Chatting and cheating in the chatbot era also include such guidelines. It is also important to be aware of different levels that can be used while developing chatbot guidelines: King’s College guidance on GenAI for teaching, assessment and feedback “aims to support the adoption and integration of generative AI at different institutional levels - macro (university), meso (department, programme, module), and micro (individual lecturers, especially those with assessment roles).” ([King’s guidance on generative AI for teaching, assessment and feedback | King's College London](#)).

Table 3. Constructive alignment

Study/ Report	Country where the education institution is based	Explicitly mentioning implications for exam boards?	Based on empirical data and/or other sources	Main result
Rudolph et al. (2023)	Singapore	No	Education/GenAI studies	“The principle of constructive alignment asks us to ensure that learning objectives, learning and teaching and assessments are all constructively aligned. (Biggs & Tang, 2011)”
Unesco (2023)	-	No	Education/GenAI studies	Review all forms of assessment and evaluation to ensure that each element is fit for purpose. This review may lead to replacing exams or other assessments with in-person assessments or altering the types of questions or exam formats that are used. (p.13)

Rudolph et al. (2023) and Unesco (2023) are right to underline the importance of constructive alignment to decide if chatbots can be used students for their assignment. Exam boards should suggest their education directors to apply a policy of constructive alignment to take end terms and learning goals as a point of departure, meaning that they decide of chatbots can be used for an assignment or not. For traditional knowledge testing chatbot use can not be allowed as examiners would not be able to assess the students' knowledge or ability correctly and fairly which means that in that case for example a take home essay as assignment is not possible anymore (see also Marche (2022) in Chapter 2). And if chatbots can be used it is important that the students are transparent about how they made use of these tools (for example by including the

prompts they used and the answers from the chatbot). Another important point is raised by Weng et al. (2024) who argue that “Educators should ensure that assessment tasks address relevant learning outcomes for each subject (van der Veen & van Oers, 2017). Assessment designs should engage students with tasks that require critical thinking and cannot be easily replicated by LLMs (Crawford et al., 2023; Kuhn, 2019; Jordanou et al., 2019).” (p.7). See also [Considering generative artificial intelligence \(GenAI\) in assessment design - Teaching & Learning](#).

And finally, as already mentioned in chapter 1, [Generative AI in education and the implications for assessment - Educational Development & Training - Utrecht University](#) and [Generative Artificial Intelligence in Learning and Teaching - Southern Cross University](#)) for more information about how to put constructive alignment into practice.

Table 4. *Training for students and lecturers*

Study/ Report	Country where the higher education institution is based	Explicitly mentioning implications for exam boards?	Based on empirical data and/or other sources	Main result
Fengchun & Holmes (2023/ 2024)	-	No	Education/GenAI studies	“Provide guidance and training to researchers, teachers and learners about GenAI tools to ensure that they understand the ethical issues such as biases in data labelling and algorithms, and that they comply with the appropriate regulations on data privacy and intellectual property.” (p. 28)

Reyna, (2023)	Australia	No	Education/GenAI studies	<p>“Faculty training is a crucial step in successfully integrating ChatGPT into higher education. Institutions should provide comprehensive professional development sessions to train educators and professional staff about the capabilities and limitations of ChatGPT. These training sessions should also emphasise the ethical considerations surrounding the use of AI technology and best practices for its integration in the classroom. By equipping faculty with the necessary knowledge and skills, they can confidently leverage ChatGPT to enhance teaching and learning experiences for their students.</p>
				<p>Establishing a community of practice (CoP) (Baker & Beames, 2016) can significantly support the effective implementation of ChatGPT in higher education. This community would bring together educators who are using or interested in using ChatGPT, allowing them to share their experiences, strategies, and resources. Through regular meetings, workshops, webinars and online platforms, educators can engage in discussions, collaborate on innovative ideas, and learn from one another’s successes and challenges. The CoP promotes a culture of continuous improvement and provides a supportive network where educators can explore the potential of ChatGPT in higher education.</p>
				<p>To ensure a smooth integration process, technical support should be readily available to address any challenges or concerns related to the implementation of ChatGPT. Higher Education institutions should offer technical assistance, troubleshooting guidance, and resources to support educators in effectively utilising the AI model. This support can range from addressing technical issues with the technology to helping educators navigate ethical considerations and privacy concerns. By providing robust technical support, institutions can alleviate concerns and enable educators to integrate ChatGPT into their teaching practices confidently.”</p>

Unesco
(2023)

Education/GenAI studies

“Training for staff can ensure that the support they provide to students and other stakeholders builds on rather than replicating what chatbots/AI tools offer and increase confidence in the deployment of technology. Peer support and mentoring for faculty members to increase skill level and share good practices for teaching and ways of using ChatGPT in research can be done within faculties, at institutional level, or among supra-institutional communities of knowledge.” (p. 13)

“Peer support and mentoring for faculty members to increase skill level and share good practices for teaching and ways of using ChatGPT in research can be done within faculties, at institutional level, or among supra-institutional communities of knowledge (p.13).

The reports and studies make clear that training for not only students but also lecturers is important for responsible use of chatbots in higher education institutes. See also the open access tutorials from Utrecht University’s Graduate School of Life Sciences [Teachers' Guide | Graduate \(School of Life Sciences | Generative AI guidelines\)](#), Utrecht University’s Centre for Digital Humanities’ trainings <https://cdh.uu.nl/training/> and the [\(The AI Maturity in Education Scan \(AIMES\) - Vrije Universiteit Amsterdam\)](#) developed by VU Amsterdam and UvA.

Table 5. Monitoring and validating GenAI systems for education

Study/ Report	Country where the higher institution is based	Explicitly mentioning implications for exam boards?	Based on empirical data and/or other sources	Main result
Fengchun & Holmes (2023/ 2024)	-	No	Education/GenAI studies	<p>(...) the development and deployment of GenAI should be ethical by design. Subsequently, once the GenAI is in use, and throughout its lifecycle, it needs to be carefully monitored and validated – for its ethical risks, its pedagogical appropriateness and rigour, and its impact on students, teachers and classroom/school relationships. In this respect, the following five actions are recommended:</p> <ul style="list-style-type: none"> ● Build validation mechanisms to test whether GenAI systems used in education and research are free of biases, especially gender biases, and whether they are trained on data representative of diversity (in terms of gender, disability, social and economic status, ethnic and cultural background, and geographic location). ● Address the complex issue of informed consent, particularly in contexts where children or other vulnerable learners are not capable of giving genuinely informed consent. ● Audit whether outputs of GenAI include deepfake images, fake (inaccurate or false) news, or hate speech. If the GenAI is found to be generating inappropriate content, institutions and educators should be willing and able to take swift and robust action to mitigate or eliminate the problem. ● Exercise strict ethical validation of GenAI applications before they are officially adopted in educational or research institutions (i.e. adopt an ethics-by-design approach). ● Before making decisions on institutional adoption, ensure that the GenAI applications in question do no predictable harm to students, are educationally effective and valid for the ages and abilities of the target learners, and are aligned with sound pedagogical principles (i.e. based on the relevant domains of knowledge and the expected learning outcomes and development of values). (p.25)

	-			“Institutional auditing of GenAI algorithms, data and outputs: Implement mechanisms to monitor as best as possible the algorithms and data used by GenAI tools and the outputs they generate. This should include regular audits and assessments, the protection of user data, and automatically filtering out inappropriate content.” (p.23)
Unesco (2023)	-	No	Education/GenAI studies	“Monitor performance and equity How effective is the AI technology in meeting the need that was identified? What criteria are used to measure effectiveness? Can the data collected be used by the HEI? How can it be used? How often is data collected? To what extent is the AI technology overcoming or addressing equity concerns? How is this measured?” (p.14)

The two reports make clear that responsible use of chatbots is not a one time check. This tells exam boards that they should work together with their education directors to embed the monitoring and validating GenAI systems for education in their institution. Utrecht University’s USO project AI in Higher Education “seeks to strengthen reflective awareness among staff and students about the risks and possibilities of generative AI. For this, educational interventions will be set in place to discover risks and possibilities. These interventions are intended to generate a representative body of findings about the risks and possibilities of using generative AI in educational contexts, and will hopefully provide us with enough input to produce informative guidelines, instructions and strategies for generative AI use.” (<https://teaching-and-learning-collection.sites.uu.nl/project/ai-in-higher-education/>).

4. Conclusion and recommendations for exam boards



4.1 Conclusion

Detecting if students used chatbots to generate texts appeared to be not possible. It is better that exam boards stimulate focusing on future ready digital media literacy skills in their curricula making students aware of the Strengths, Weaknesses, Opportunities and Threats of chatbots. In the next section we formulate recommendations to reach this goal. It should also be noted that the narrative literature (in Tables 1- 5, Chapter 3) showed that *none of the papers* explicitly addressed the point of responsible chatbot use for university students *from the perspective of exam boards*, but we found papers addressing this topic *for lecturers and students*, allowing us to draw conclusions from the insights they offer for exam boards. These studies are useful but we also showed that in most cases the results are not based on empirical studies but on other sources such as reports by Unesco and the Australian Government, a content analysis from newspapers discussing the impact of chatbots, an evaluation of popular plagiarism detection tools to evaluate the originality of 50 essays generated by ChatGPT on various topics and the measurement of the attitudes toward aigiarism and plagiarism.

4.2 Recommendations for exam boards

For all recommendations that will finally follow now it is important for exam boards to collaborate with their education directors, lecturers and students.

1. Be aware that detecting if a text is generated by a chatbot is impossible (see Chapter 2).
- 2a. Develop chatbot guidelines to avoid fraud - Be aware that if the exam committee punishes a student who admits s/he used a chatbot in a non responsible way, but does not punish a student who denies this, this risks to lead to legal inequality. Therefore, consider to follow the recommendations below.
- 2b. It is important to develop chatbot guidelines in collaboration with your education directors, students and lecturers.

3. Stimulate actions at micro, meso en macro level in your institute (see Table 2).
4. Develop chatbot guidelines for lecturers focusing on end terms and learning goals by following the principle of constructive alignment (see  2., Table 3 and [Generative AI in education and the implications for assessment - Educational Development & Training - Utrecht University](#)) while redesigning assessments (see Chapter 3).
5. It is important to decide in which cases its use should be limited to the minimum, allowed with restrictions, or propagated, following the principle of constructive alignment (see section 3.1). Let lecturers test different kinds of assignments, with and without chatbot use. See also Utrecht University's USO project AI in Higher Education (<https://teaching-and-learning-collection.sites.uu.nl/project/ai-in-higher-education/>)
6. Contact other exam boards in- and outside your university for the GenAI guidelines they developed (see  **Boxes 1, 2 and 3**).
- 7a. Train not only students but also lecturers about responsible chatbot use (see Table 4).
- 7b. Get inspired by the trainings offered by Utrecht University's Graduate School of Life Sciences ([Teachers' Guide | Graduate School of Life Sciences | Generative AI guidelines](#)) and Utrecht University's Centre for Digital Humanities' trainings (<https://cdh.uu.nl/training/>).
8. Stimulate peer support and mentoring for lecturers (see Table 4).
9. Take the initiative to monitor and validate GenAI systems for education. (see Table 5).
10. Take the initiative to organize an AI audit in higher education (see Table 5).

We would like to end this research report with the following quote:

“Higher education reactions to ChatGPT and GPT-3 have been on a continuum between the extremes of banning or prohibiting the use of the software and including it in the curricula. How should students, teachers and higher education institutions deal with ChatGPT? Marche (2022) predicted that it may take “10 years for academia to face this new reality: two years for the students to figure out the tech, three more years for the professors to recognize that students are using the tech, and then five years for university administrators to decide what, if anything, to do about it.” Rudolph et al. (2023, p.13)

It may take our universities 10 years (or more or less?) to fully integrate chatbots in their educational practices, but exam boards should continue on the way they started: further developing a framework based on guidelines that can be used by lecturers to train their students to use chatbots in a responsible way. This allows them not only to assess the students' knowledge or ability correctly and fairly but also to provide them with “future ready digital media literacy skills” making them aware of the Strengths, Weaknesses, Opportunities and Threats of this digital tool.

References

- Anders, B. A. (2023). Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking? *Patterns*, 4(3). <https://doi.org/10.1016/j.patter.2023.100694>
- Baker, A., & Beames, S. (2016). Good CoP: What Makes a Community of Practice Successful? *Journal of Learning Design*, 9(1), 72-79. <https://doi.org/10.5204/jld.v9i1.234>
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). McGrawHill. ISBN: 978-0335242757
- Cassidy, C. (2023, January 10). Australian universities to return to 'pen and paper' exams after students caught using AI to write essays. *The Guardian*. <https://www.theguardian.com/australia-news/2023/jan/10/universities-to-return-to-pen-and-paper-exams-after-students-caught-using-ai-to-write-essays>
- Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in education and teaching international*, 61(2), 228-239. <https://doi.org/10.1080/14703297.2023.2190148>
- Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice*, 20(3), 2. <https://doi.org/10.53761/1.20.3.02>
- Deutsche Welle (2023). AI experts say ChatGPT is changing education. But how? <https://frontline.thehindu.com/news/ai-experts-say-chatgpt-is-changing-education-but-how/article66449967.ece>
- Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press. ISBN: 978-1452242972
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861-874. <http://doi.org/10.1016/j.patrec.2005.10.010>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460-474. <https://doi.org/10.1080/14703297.2023.2195846>
- Fengchun, M., & Holmes, W. (2023/2024). Guidance for generative AI in education and research. Education 2030. United Nations Educational, Scientific and Cultural Organization. [7 September 2023, last update:19 June 2024] <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
- Ferrari, F., van Dijck, J., & van den Bosch, A. (2023). Foundation models and the privatization of public knowledge. *Nature Machine Intelligence*, 1-3. <https://doi.org/10.1038/s42256-023-00695-5>

- Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, 15(2), ep421 <http://doi.org/10.30935/cedtech/13036>
- Iordanou, K., Kuhn, D., Matos, F., Shi, Y., & Hemberger, L. (2019). *Learning by arguing*. *Learning and Instruction*, 63, 101207. <https://doi.org/10.1016/j.learninstruc.2019.05.004>
- Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez-Maldonado, R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence*, 8, 100348. [10.48550/arXiv.2405.11800](https://doi.org/10.48550/arXiv.2405.11800)
- Keith, T. (2022, February 16). Combating academic dishonesty, Part 1 – understanding the problem. The University of Chicago. <https://academictech.uchicago.edu/2022/02/16/combating-academic-dishonesty-part-1-understanding-the-problem/>
- Khalaf, M. A. (2024). Does attitude towards plagiarism predict aigiarism using ChatGPT? *AI and Ethics*, 1-12. <https://doi.org/10.1007/s43681-024-00426-5>
- Khalil, M., & Er, E. (2023, June). Will ChatGPT G et You Caught? Rethinking of Plagiarism Detection. In *International Conference on Human-Computer Interaction* (pp. 475-487). Springer Nature Switzerland. http://doi.org/10.1007/978-3-031-34411-4_32
- Kim, S. (2022, December 11). ChatGPT Output How to detect if the student used OpenAI's ChatGPT to complete an assignment. *Geek Culture*. <https://medium.com/geekculture/how-to-detect-if-an-essay-was-generated-by-openais-chatgpt-58bb8adc8461>
- King, M. R., & ChatGPT. (2023). A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cellular and molecular bioengineering*, 16(1), 1-2. <https://doi.org/10.1007/s12195-022-00754-8>
- Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability*, 15(7), 5614. <https://doi.org/10.3390/su15075614>
- Kuhn, D. (2019). Critical thinking as discourse. *Human Development*, 62(3), 146-164. <https://doi.org/10.1159/000500171>
- Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's Research Reports*, 6(1).
- Lancaster, T. (2023). Artificial intelligence, text generation tools and ChatGPT—does digital watermarking offer a solution? *International Journal for Educational Integrity*, 19(1), 10. <https://doi.org/10.1007/s40979-023-00131-6>
- Liu, D., & Bridgeman, A. (2023, July 12). What to do about assessments if we can't out-design or out-run AI? [What to do about assessments if we can't out-design or out-run AI? – Teaching@Sydney](https://teaching@sydney.edu.au/2023/07/12/what-to-do-about-assessments-if-we-cant-out-design-or-out-run-ai/)

- Lodge, J., Howard, S., & Bearman, M. (2023). *Assessment reform for the age of artificial intelligence*. Australian Government, Tertiary Education Quality and Standard Agency. <https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/assessment-reform-age-artificial-intelligence>
- Loos, E., Gröpler, J., & Goudeau, M. L. S. (2023). Using ChatGPT in Education: Human Reflection on ChatGPT's Self-Reflection. *Societies*, 13(8), 196. <https://doi.org/10.3390/soc13080196>
- Loos, E., & Radicke, J. (2024). Using ChatGPT-3 as a writing tool: an educational assistant or a moral hazard? Current ChatGPT-3 media representations compared to Plato's critical stance on writing in Phaedrus. *AI and Ethics*, 1-14. <https://doi.org/10.1007/s43681-024-00470-1>
- Mai, D. T. T., Da, C. V., & Hanh, N. V. (2024, February 9). The use of ChatGPT in teaching and learning: a systematic review through SWOT analysis approach. *Frontiers in Education*, 9, p. 1328769. <https://doi.org/10.3389/educ.2024.1328769>
- Marche, S. (2022, December 7). The college essay is dead. Nobody is prepared for how AI will transform academia. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/>
- McDonald, N., Johri, A., Ali, A., & Collier, A. H. (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans*, 100121. <https://doi.org/10.1016/j.chbah.2025.100121>
- Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, 5, 100151. <https://doi.org/10.1016/j.caeo.2023.100151>
- Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching*, 6(1), 41-56. <https://doi.org/10.37074/jalt.2023.6.1.29>
- Reyna, J. (2023). *AI in the Classroom: A Comprehensive Framework for ChatGPT Integration in Teaching and Learning in Higher Education* (1-14). https://www.researchgate.net/publication/371489410_AI_in_the_Classroom_A_Comprehensive_Framework_for_ChatGPT_Integration_in_Teaching_and_Learning_in_Higher_Education
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of applied learning and teaching*, 6(1), 1-22. <https://doi.org/10.37074/jalt.2023.6.1.9>
- SEJ SEO (2023, January, 11). How The ChatGPT Watermark Works And Why It Could Be Defeated. *Search Engine Journal*. <https://www.searchenginejournal.com/chatgpt-watermark/475366/>

- Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching*, 6(1). <https://doi.org/10.37074/jalt.2023.6.1.17>
- Unesco (2023). ChatGPT and artificial intelligence in higher education: quick start guide. United Nations Educational, Scientific and Cultural Organization. Published April 2023, based on ChatGPT-3 <https://unesdoc.unesco.org/ark:/48223/pf0000385146>
- van Bergen, P., & Lane, R. (2016, November 30). Should we do away with exams altogether? No, but we need to rethink their design and purpose. *The Conversation*. <https://theconversation.com/should-we-do-away-with-exams-altogether-no-but-we-need-to-rethink-their-design-and-purpose-67647>
- van der Veen, C., & van Oers, B. (2017). Advances in research on classroom dialogue: Learning outcomes and assessments. *Learning and Instruction*, 48, 1-4. <https://doi.org/10.1016/j.learninstruc.2017.04.002>
- van Es, K., Schäfer, M. T., & Wieringa, M. (2021). Tool Criticism and the computational turn. A “methodological moment” in media and communication studies. *M&K Medien & Kommunikationswissenschaft*, 69(1), 46-64. <https://doi.org/10.5771/1615-634X-2021-1-46>
- Wang, H., Dang, A., Wu, Z., & Mac, S. (2024). Generative AI in higher education: Seeing ChatGPT through universities' policies, resources, and guidelines. *Computers and Education: Artificial Intelligence*, 7, 100326. <https://doi.org/10.1016/j.caeai.2024.100326>
- Weng, X., Qi, X. I. A., Gu, M., Rajaram, K., & Chiu, T. K. (2024). Assessment and learning outcomes for generative AI in higher education: A scoping review on current research status and trends. *Australasian Journal of Educational Technology*, 40(6). <https://doi.org/10.14742/ajet.9540>
- Wineburg, S., & McGrew, S. (2016). *Evaluating information: The cornerstone of civic online reasoning*. Stanford History Education Group. <https://purl.stanford.edu/fv751yt5934>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Williamson, B., Molnar, A., & Boninger, F. (2024). *Time for a Pause: Without Effective Public Oversight, AI in Schools Will Do More Harm Than Good*. University of Edinburgh, National Education Policy Center (NEPC).
- Zhang, H., Edelman, B. L., Francati, D., Venturi, D., Ateniese, G., & Barak, B. (2023). Watermarks in the sand: Impossibility of strong watermarking for generative models. *arXiv preprint arXiv:2311.04378* <https://doi.org/10.48550/arXiv.2311.04378>

Appendix 1

Guidelines for detection of misuse of GenAI (original in Dutch)

[Internal document from Utrecht University, Faculty of Humanities, Taskforce Responsible and Ethical Use of GenAI - original version in Dutch, translated DeepL.com (free version), 11.01.2025]

1. How can potential misuse of GenAI be fairly detected?
2. What to do when abuse is suspected?

How to fairly detect potential misuse of GenAI?

The unauthorised use of AI is difficult to prove, but it is certainly not impossible. Existing detection software for AI use is unreliable and not suitable for demonstrating fraud demonstrate. To discuss possible misuse of GenAI with a student, it is important that they are aware aware of the guidelines for its use in your course. Ensure that these guidelines are included in the course manual and that you discuss them when giving instructions for assignments.

What to do before the course starts:

- Put your own assignment instructions into GenAI and see what the result looks like. Look specifically at the quotes in the text, style and language use (flat, generic), structure, sweeping statements.
- Agree with the study group tutors what the writing assignments might look like. Take for example, a number of previously written texts (weak and strong examples) and a GenAI-written text, and compare them in relation to the learning objectives.

What to do during the course:

- Collect draft versions and monitor the process.
- Check the final products for possible misuse of GenAI.

Some indications you can look out for (depending on what is allowed in the course):

- The work differs significantly from the student's previously submitted work
- Statements that deviate significantly from the course material
- Unexplained deviations in text length of instructions
- Discrepancies in language use
- Clear structure (beginning, middle, end; 1,2,3) but content of the text is flat and general
- Lack of own opinion or reflection (GenAI is good at imitation, not originality)
- Inconsistent tone or style
- Sweeping statements
- Quotations:
 - o Lack of citations
 - o Quotes that do not match course content
 - o Incorrect quotes

What should you do if you suspect abuse?

If you suspect that a student has used AI against the guidelines of the course manual, invite that person to a meeting to discuss the matter. Be sure not to accuse the student directly. Raise your concerns about their work based on the indications above. During the conversation, ask the student to explain their writing process and reasoning to explain.

- How did you develop your idea?
- What methods did you use to research the topic?
- What sources did you consult?
- How did you write this paper?
- How much time did you spend completing the assignment?
- How long did the editing take?

If the student cannot clear your doubts during the interview and the suspicion of fraud persists, report the matter to the course coordinator and the examination board. For the examination board, it is important that you make the basis of your suspicions (based on the indicators and the conversation with the student) clearly motivate.

References

Algonquin College. 2023. "Faculty Guide: Fair Detection of GenAI Misuse." Laatst gewijzigd

November 28, 2023. [https://www.algonquincollege.com/academic-integrity/files/2023/11/Faculty Guide-Fair-Detection-of-GenAI-Misuse-Nov28.2023.pdf](https://www.algonquincollege.com/academic-integrity/files/2023/11/Faculty%20Guide-Fair-Detection-of-GenAI-Misuse-Nov28.2023.pdf).

Other sources

- Fraud and plagiarism - Intranet (uu.nl)

- Concordia University. 2024. "Generative AI Teaching Guidelines." Laatst gewijzigd, juli 2024. <https://www.concordia.ca/ctl/tech-tools/teach-with-technology/guidelines-gen-ai.html#misuse>

- University of Manitoba. 2024. "Artificial Intelligence: How Can I Tell If a Student Used a GenAI System on My Assessment?" Laatst gewijzigd, 2024. <https://umanitoba.ca/centre-advancement-teaching-learning/integrity/artificial-intelligence#how-can-i-tell-if-a-student-used-a-genai-system-on-my-assessment>

- University of Amsterdam. 2024. "Unsupervised Written Assignments." Laatst gewijzigd, 2024. <https://tlc.uva.nl/en/article/unsupervised-written-assignments>

- University of Saskatchewan. 2023. "How to Talk with Students About Suspected Academic Misconduct." Laatst gewijzigd, 23 oktober 2023. <https://sites.usask.ca/gmcte/2023/10/23/how-to-talk-with-students-about-suspected-academic-misconduct>

Appendix 2 GenAI detectors

[Parts from an internal document from Utrecht University by Van den Bosch, 10.03.2024, original version in Dutch translated DeepL.com (free version), 11.01.2025]

1. <https://openai-openai-detector.hf.space/>, the GPT-2 detector - this detector was published 4 years ago by OpenAI, at the same time as GPT-2. The open source software, which provides an online interface at the aforementioned URL into which texts can be pasted, 1 Most detectors report on AUC, Area under the ROC Curve; see

<https://www.sciencedirect.com/science/article/abs/pii/S016786550500303X>. This metric is based on two components, false positive rate and true positive rate (recall); it is usually only reported on AUC and not on the underlying components, even though that is relevant in this context. Detects whether the text was generated by GPT-2, a predecessor of the later versions of GPT as they are now in ChatGPT (GPT-3.5) or GPT-4. According to the summary report released on GPT-2 and the detector, the detector is 'able to detect 1.5 billion parameter GPT-2-generated text with approximately 95% accuracy'. The report warns that the detection task becomes more difficult with larger models. GPT-3.5 and GPT4 are more than 100 times larger than GPT-2 in numbers of parameters. The detector has not been trained or tested on GPT-3 and is therefore not useful in advance as a detector of texts generated by ChatGPT (based on a version of GPT-3).

2. <https://platform.openai.com/ai-text-classifier>, the GPT-3 detector - this detector was published by OpenAI on 31 January 2023, with no underlying open source code. The classifier classifies an input text into one of five categories: Very unlikely to be AI-generated, Unlikely to be AI-generated, Unclear if it is AI written, Possibly AI-generated, Likely AI-generated. In the information provided on the page of this interface, we read:

'We evaluated our classifier as well as our previously published classifier [the GPT-2-detector-AB] on a held-out validation set with the same distribution as the training set ('validation set') and on a challenge-set composed of human-written completions and completions from a strong language model trained on the human completions ('challenge set'). We find that our classifier outperforms our previously published classifier, with an AUC score of 0.97 on the validation set and 0.66 on the challenge set (compared to 0.95 on the validation and 0.43 on the challenge set for our previously published classifier). We also find that the classifier's performance degrades as the size of the model that is generating text increases. In other words, as the size of a language model increases, its outputs look more like human-written text to the AI text classifier.'

3. <https://gptzero.me/>, GPT-Zero - this detector launched in January 2023 appears to be fairly accurate with an AUC of 0.98 on its own (non-public) validation data, but does indicate in <https://gptzero.me/faq> that it is less good at detecting AI-generated text than human-generated text. GPT-Zero seems to be able to detect ChatGPT / GPT-3-generated text, but perhaps also other AI text generators.

4. <https://writer.com/ai-content-detector/> - a detector for all types of AI-generated text, not reporting on scores on validation material.

5. <https://copyleaks.com/ai-content-detector> - a detector for all types of AI-generated text. An accuracy of 99.2% is claimed, with no further background information.

6. <https://detectgpt.ericmitchell.ai/> - a detector of GPT-2 output. This detector is published only as a demo and proof of concept, with no claims.