

An attempt to cheat using GPT-4: findings, discussion and recommendations for academic staff and students

Nick Mapletoft *, University Centre Quayside (UCQ), Newcastle upon Tyne, UK, nick.mapletoft@ucq.ac.uk,

Andy Price, UCQ, Newcastle upon Tyne, UK, andy.price@ucq.ac.uk

Kevin Smith, UCQ, Newcastle upon Tyne, UK, kevin.smith@ucq.ac.uk

Olga Mapletoft, UCQ, Newcastle upon Tyne, UK, olga.mapletoft@ucq.ac.uk

Michelle Elliott, UCQ, Newcastle upon Tyne, UK, michelle.elliott@ucq.ac.uk

Abstract

This manuscript explores the use and detection of ChatGPT artificial intelligence (AI) content, in the academic setting. The study involves an academic staff and student partnership at a Higher Education Provider (HEP), examining the implications of the chatbot's use in academic writing and its potential impact on educational practice.

We employ a dialogical approach in this Scholarship of Teaching and Learning (SoTL) study, mirroring the ChatGPT conversational style. A former student uses ChatGPT-4 to rewrite an essay and the generated content is evaluated by free-to-use AI checking tools, Turnitin, and the module tutor.

Our findings illustrate that while ChatGPT-4 could generate a passable assignment, it lacks depth, criticality, and contextualisation. Further, AI checking tools, while useful, exhibit limitations. However, the generated content could serve as a starting point for students, suggesting a possible shift in the educational paradigm with AI's integration.

The manuscript also underscores the importance of understanding and navigating AI responsibly, calling for clear guidelines for AI use and improved methods of detecting AI generated content. While AI has emerged as an efficient tool for generating academic content, our study highlights the necessity of human-centric learning and critical thinking skills in academia, emphasising that AI should not replace personal research. We also discuss potential changes to assessment methods in response to AI advancements,

* Corresponding Author
ISSN 2755-9475 (Online)



This is an Open Access article distributed under the terms of the [Creative Commons –Attribution License 4.0 International](https://creativecommons.org/licenses/by/4.0/) which permits use, distribution, and reproduction in any medium, even commercially, provided the original work is properly attributed.

recommending strategies for both academic staff and students. Our findings have implications for educational institutions worldwide as AI tools become increasingly accessible, raising questions about academic integrity, pedagogical strategies, and the evolving relationship between technology and education.

Keywords: ChatGPT, GPT-4, AI, Artificial Intelligence, Turnitin, LLM, academic integrity

Background

ChatGPT is an AI chatbot developed by OpenAI, an organisation founded in 2015 with a mission to develop safe generative AI that can interact with humans in a conversational manner (Taecharungroj, 2023). The third-generation generative pre-trained transformer (GPT-3) (Dale, 2021) is a large language model (LLM) that was publicly released on 30th November 2022, and immediately took off in record-breaking fashion (Duarte, 2023). In 2023 this was followed by GPT-4, which is said to be able to pass examinations (Ali et al., 2023; Katz et al., 2023; Lin et al., 2023). This meteoric rise in popularity has led many staff working in HEPs to not know how they should react; opinion is divided over whether the use of ChatGPT should be restricted or even legislated (Dwivedi et al., 2023).

The integration of ChatGPT and similar AI tools in higher education has spurred diverse perspectives from stakeholders including educators, students, administrators and information technology practitioners. The debate is multifaceted and teeming with nuances. Proponents argue that these AI tools can revolutionise the academic landscape by offering instant, personalised feedback tailored to individual learning needs, especially beneficial in resource-constrained environments or large class settings. This potential democratisation of knowledge could bridge gaps, providing unparalleled access to those who might otherwise remain voiceless in some academic settings. There is clearly a role for AI in business and the Higher Education Provider's (HEP's) role involves equipping students for their employment, so AI cannot simply be excluded from curricula. Additionally, the lure of potential institutional cost savings and the tool's capability to monitor and deter academic dishonesty further bolster the case for its inclusion. However, this enthusiasm is met with equally compelling reservations. Detractors caution against the risk of engendering a superficial grasp of subjects, as students might seek shortcuts, potentially leveraging the AI to sidestep genuine engagement with their studies.

AI tools are sophisticated, however some of the generated content is inaccurate, misleading, and in some cases racially biased (Bender et al., 2021; Lindebaum, 2023). Furthermore, while the technology could democratize access, it could paradoxically deepen the digital divide, leaving behind those without the requisite technological access. Additionally, whilst AI's surveillance capabilities could combat cheating, there's a lurking concern about students using these tools unethically, eroding academic values. Data privacy looms large, with apprehensions about the security and ethical implications of gathering and processing vast student data sets and questions about data ownership. The traditional role of educators is also under the microscope; while some see AI liberating educators from mundane tasks including marking (Huawei & Aryadoust, 2023), allowing a focus on deeper pedagogical endeavours, others fear a diminishing human touch in education, possibly even job displacements. This drift towards possible depersonalisation, where education feels increasingly mechanistic, is unsettling for many. At its core, beyond the tangible advantages or pitfalls, lies an even profounder philosophical conundrum about AI's ethical place in moulding human thought, values, and our very cultural fabric. As higher education grapples with these complexities, striking the right balance becomes pivotal, ensuring that technology amplifies academic virtues without overshadowing them.

Introduction

This manuscript involves a SoTL approach, through an academic staff / student partnership at an HEP, undertaking systematic enquiry into the use and detection of ChatGPT generated content, and then sharing the results publicly. One of the authors is the Module Tutor and marker for 'New Technologies'. Two of the authors, both member of the HEP's Senior Leadership Team (SLT), previously became students on the HEP's programme, to experience the programme as students and to be able to advise and provide a former student's perspective. The former students undertook the 'New Technologies' module in 2018, prior to the widespread adoption of AI LLM tools. This manuscript involves one former student using ChatGPT-4 (GPT-4) in 2023 to rewrite their essay. The paper describes the GPT-4 conversation, and the AI-generated assignment submission is then checked by popular free-to-use AI checking tools, Turnitin and the Module Lead who is the assignment marker. This research was undertaken to provide

utilitarian support to academic staff and students, resulting from a first-hand account, with a view to illuminating the use and detection of ChatGPT in a real situation.

The contributors decided to take a dialogical approach to this research and manuscript, to best reflect the benefits of dialogic teaching (Kim & Wilkinson, 2019) involving the ongoing conversation between the tutor and students to mirror the conversational approach used when interacting with ChatGPT. Dialogic teaching enables students to explore their own understanding and at the same time, to practice the use of language as a knowledge construction tool (University of Cambridge, 2023) This dialogic approach also aligns with SoTL's focus on ongoing interactions and conversation between educators and students, to improve learning and teaching.

University Centre Quayside (UCQ) is an HEP delivering its business and management bachelor's degree in England. This manuscript has a much wider geographical dimension because ChatGPT is accessible globally, so this work will naturally be of interest to an international audience. India, Japan and Brazil are the biggest users of ChatGPT after the United States of America (Duarte, 2023). Collaborating globally in 'real time' will become the new normal for HE organisations as they try to stay up-to-date with teaching and learning practice and policy development. There are established international conferences that address this issue, for example, the [European Conference on Ethics and Integrity in Academia 2023](#).

The student's conversation with GPT4 to generate an alternative essay, in the student's own words

I had a conversation with GPT4 to generate an alternative version of my essay, the full conversation is available [here](#). My first question was taken directly from the assignment brief in the module guide "can you provide an evaluation of the impact of current and predicted future digital technologies on organisations?". The response provided details of six key impacts, each with its own separate paragraph, consisting of 'Operational Efficiency', 'Business Model Innovation', 'Customer Experience', 'Remote Work', 'Cybersecurity' and 'Sustainability'. There were also four future predictions, and a paragraph was provided for each of 'Quantum Computing', 'Metaverse', 'AI and Ethics' and '5G and Beyond', followed by a conclusionary paragraph.

The selection of six key impacts and four future predictions was considered as an appropriate response to the question and could make a good basis for the essay. The issue was that the corresponding text GPT4 generated was shallow, lacking any depth of criticality. Teaching critical thinking is notoriously challenging (Willingham, 2007), however, by this stage students would be expected to write critically, in order to demonstrate that they can think critically (Coe, 2011). There are of course different types of critical thinking. Barnett (2022, pp. 142–143) provides us with a list of twelve different types, which gives an indication of some of the types that a student might incorporate in this essay. For example, being able to logically interrogate texts to ascertain their validity; evaluating text against a set of standards, such as in the student's own discipline; to evaluate a text against a wider cognitive framework; via critical self-reflection, assessing the student's own work in a given context; to demonstrate authenticity by forming their own ideas. The GPT4-generated response did not demonstrate any such criticality.

The initial response was also without any academic references, so I asked, "Can you regenerate with academic references? GPT-4 responded by rewriting the same six key impacts and four future predictions, but this time with in-text citations and an accompanying reference list. On the face of it, what GPT-4 had produced, looked like a typical academic essay and it appeared to be answering the question. For example, the GPT-4 output for its first suggested key impact was:

"Operational Efficiency: Digital technologies like cloud computing, automation, artificial intelligence (AI), and machine learning (ML) have facilitated operational process improvements, enabling rapid data analysis, reducing human error, and boosting efficiency (Bughin et al., 2017). Advanced algorithms and predictive analytics provide insights for informed decision-making (Provost & Fawcett, 2013)." (OpenAI, 2023)

On initial observation the GPT4 citations and references appear to be accurately and consistently presented. Many of the references were valid and accurate; Ahshami et al.'s, (2023) quantitative study determined that ChatGPT's overall accuracy for article discarding and classification was 88%. On further checking however some of the references were found to be inaccurate and misleading in substance, with some being completely fabricated. For example, GPT4 presented this paper: "Massingham, P. (2019). Does Leadership Promote Knowledge Management?. Knowledge and Process

Management, 26(3), 232-242.” This appears to be a plausible reference, but it is not. Peter Massingham has written about knowledge management but does not appear to have written a paper with this title; no paper with that title could be found. Checking the Knowledge and Process Management journal, Volume 26, Issue 3 was indeed published in 2019, but there was not an article by Massingham and the pages referenced were incorrect as pp 229-243 was an article by others (Curado et al., 2019). Peter Massingham has co-authored an earlier article that was published in this journal entitled “Embedding Knowledge Management into Business Processes” (Massingham & Al Halaibi, 2017). Other references, whilst accurate, for example that by Jennex and Olfman (Jennex & Olfman, 2006), were neither contextually pertinent nor in date; cited sources for this assignment should focus on the last five years (with a caveat that peer reviewed journal articles and book chapters may have been written a year or two before being published). None of the on articles on the module’s suggested reading list were provided by GPT4. These findings concur with a systematic review of ChatGPT undertaken by Singh and Singh (2023, pp. 193–212) which determined that, whilst ChatGPT generated content is remarkable, it is not always good.

Whereas the first assignment question was generic, the second question specifically relates to the student’s own organisation. In this case, my organisation is the University Centre Quayside (UCQ). I asked the question, “Can you [provide] An evaluation of knowledge management culture within UCQ?” To which GPT-4 responded with:

“As an AI developed by OpenAI, I don't have access to real-time or specific institutional data, so I cannot provide an accurate evaluation of the knowledge management culture within University College of Qatar (UCQ) or any other specific institution. This is due to the privacy constraints and data access limitations set to protect proprietary or confidential information.” (OpenAI, 2023). However, it then went on to provide a non-institutionally specific response, without in-text citations or references. So I then asked, “Can you fabricate an answer based on UCQ as an English University Centre and include academic references?”, to which GPT-4 responded with:

“While I can't provide a direct assessment of University College of Qatar's (UCQ) knowledge management culture, I can hypothesize based on widely accepted academic

principles of knowledge management in higher education institutions. This conjecture will be a hypothetical scenario, and it will not reflect the exact situation of UCQ. Here's an evaluation based on this presumption:

Knowledge management in educational institutions such as UCQ, which is a unique English-language center providing western-aligned tertiary education in Qatar, is integral for academic excellence, innovation, and to maintain competitiveness in the sector.” (OpenAI, 2023)

Noting that GPT-4 was referring to UCQ as University College of Qatar, and that it was lacking in-text citations and references, I asked, “Can you fabricate an answer based on UCQ as an English University Centre and include academic references?” The response still referred to University Centre of Qatar, however it was responding as if UCQ were an academic institution and its response now contained citations and references. The response provided five separate paragraphs relating to ‘Information Sharing’, ‘Leadership and Management Support’, ‘Learning Culture’, ‘Collaboration and Teamwork’ and ‘Technology’. A glance at the references showed that they were dated 2013, 2000, 2002, 2004 and 2017. So I asked, “Can you regenerate with more recent references please?” To which GPT-4 responded with “Apologies for the outdated references. Here's the revised response with more recent references:” (OpenAI, 2023) This time with references dated 2019, 2019, 2020, 2020 and 2020 respectively.

For the third assignment question I asked for “Recommended improvements to improve knowledge management culture within a university, linked to best practice.”

This was followed by six paragraphs detailing recommended improvements ‘Strengthen Leadership Commitment’, ‘Facilitate Open Communication’, ‘Leverage Technology’, ‘Provide Training and Resources’ and ‘Incentivize Knowledge Sharing’ and ‘Embed Knowledge Management in Organizational Processes’. The generated response was only 286 words, and I needed nearer to 700 so I asked, “Can you expand that answer to 700 words?” GPT-4 regenerated the response and this time it was 646 words long. Finally I asked, “Are you able to introduce criticality into that answer?” The response then made an attempt at being critical, for example:

“Strengthen Leadership Commitment: While leadership involvement is essential, it should not result in top-down decision-making that undermines the collaborative nature of

knowledge management. It's important that the commitment from leadership fosters an environment where all members of the university community feel their contributions are valued and appreciated, rather than feeling pressured to participate in initiatives they don't fully understand or agree with (Massingham, 2019)" (OpenAI, 2023). Note also that GPT-4 did not need to be told to add citations and references for section three, it did so based upon the previous requests.

Final words from the student

The GPT-4 version was produced in ten minutes and would be useful by providing an example, structure and some suggested reading; what it lacked however, was a depth of criticality. My original submission was written in about twenty hours over an eight-week period, it received a mark of 75%, compared with the GPT-4 version which the marker said would be in the 40%-49% bracket, notwithstanding that, if submitted, it would be academic misconduct. GPT-4 can be helpful as a starting point, however students must be cognisant that GPT-4 must not be used as an alternative to the student undertaking the work, the results can be shallow and the references may not exist, or may be dated.

Findings amongst various free-to-use GPT-4 detection tools

There are several tools that claim that they can detect AI- (and GPT-4-) generated content. The GPT-4 generated assignment, or parts of the assignment, depending on character or word limits, were run through the tools that were free to use. The findings are summarised in Table 1 and then explained in more detail below.

Table 1: Summary of Findings Amongst Free-To-Use GPT-4 Detection Tools

AI Detection Tool	Tool Usage Basis	In summary, this tool found that the AI-generated text was:
Writer's AI Content Detector	Free to use for 1,500 characters	"100% human-generated content"
GPTZero	Free to use for 5,000 characters	"Your text is likely to be written entirely by a human"
GPTRadar	Free to use only once	"Likely Human Generated"
ZeroGPT	Free to use and no character or word limit	"Your File Content is Most Likely Human written". Fourteen complete sentences and ten partial references were highlighted as being potentially AI generated.
Corrector App	Free to use for up to 800 words	"Fake: 0%"
OpenAI's AI Text Classifier	Free	The tool was unavailable.
Originality.ai	Paid subscription	Despite stating that it is the "Best Free Plagiarism Checker" (Jacob, 2023), this tool was not tested because it is only available through a paid subscription.

[Writer's AI Content Detector](#) is free to use, however it is limited to 1,500 characters. The first 1,324 characters of the GPT-4 generated assignment, starting from Section One, resulted in the Writer tool determining that the content was "100% human-generated content" (Writer, 2023).

Similarly, this tool determined that the original human generated assignment was "100% human-generated content" (Writer, 2023).

[GPTZero](#) is free to use for up to 5,000 characters and claims to be able to detect ChatGPT, GPT3, GPT4, Bard, human and AI + human text. It determined “Your text is likely to be written entirely by a human” amongst the first 5,000 characters. Despite having a statement saying “Sentences highlighted are more likely to be written by AI” (GPTZero, 2023), no sentences were highlighted, indicating that it did not determine that any of the analysed text has been written by AI. GPTZero analyses two variables: perplexity and burstiness. Perplexity measures the likely predictability of word strings in text, and burstiness is a measure of the complexity of the resultant sentences (Alexander et al., 2023). GPTZero assumes that AI writing has lower perplexity and burstiness. In contrast, human writing is more likely to have higher levels of burstiness and perplexity, because the human mind is significantly more complex than current AI trained tools (A. Singh, 2023). In the ‘Stats’ section, this tool gave an ‘Average Perplexity Score’ of 1628.102 and a ‘Burstiness Score’ of 8207.028, where a score of 10 is a clear indicator of AI generated text, leading the tool to state that the text was written by a human.

An analysis of the original assignment found that “There is 0% probability this text was entirely written by AI” and that “0 sentences are likely AI generated” (GPTZero, 2023).

[GPTRadar](#) was free to use, only for the first use. It determined that the GPT-4 generated assignment was “Likely Human Generated” (GPTRadar, 2023) with a 72% Accuracy and a Perplexity score of 16.

A check of the original assignment found that it was “likely Human Generated” with 81% Accuracy and a Perplexity score of 50 (GPTRadar, 2023). The perplexity was substantially higher in the human generated assignment, however the accuracy was only 9% higher. A more detailed analysis of the findings showed that this tool thought that many text chunks were AI generated.

[ZeroGPT](#) is free to use and was able to analyse the entire assignment. It claims to be “the most Advanced and Reliable ChatGPT, GPT4 & AI Content Detector”. It determined that “Your File Content is Most Likely Human written” (ZeroGPT, 2023) and gave an AI GPT score of 15.98%. Sentences and references that it thought may have been AI generated (“Highlighted text is suspected to be the most likely generated by AI”), it highlighted in yellow. In all this was fourteen complete sentences and ten partial references highlighted.

For the original assignment, ZeroGPT gave an AI GPT score of only 1.73%, highlighting just one sentence as being likely to have been AI generated and concluding that “Your Text is Human Written” (ZeroGPT, 2023).

[Corrector App](#) is free to use to check up to 800 words. All checks, tried on various computers and in different browsers, resulted in it determining “Fake: 0%” (Corrector App, 2023).

Checking the original assignment, Corrector App determined that this was “Fake: 24.07%” (Corrector App, 2023). The tool was inaccurate at analysing both the AI-generated and human-generated assignments.

[OpenAI’s AI Text Classifier](#) was unavailable, with a website message stating “As of July 20, 2023, the AI classifier is no longer available due to its low rate of accuracy” (Kirchner et al., 2023).

[Originality.ai](#) states that it is “Built for Publishers, Agencies and Writers... Not for Students” (Originality.ai, 2023). Despite being advertised as the “Best Free Plagiarism Checker” (Jacob, 2023), this is in fact a paid subscription service and so was not tested within this sample.

Findings from Turnitin

[Turnitin](#) – is not free to use, it is an institutional subscription-based service for similarity checking. Turnitin differs from most algorithmic AI detection tools in that it also has an extensive proprietary archive of work submitted to Turnitin, in addition to web harvested content, which it can use to check similarity, and to undertake AI checks. Both assignments were checked by Turnitin. The original (2018) assignment was correctly found to have 0% AI, whereas on the GPT-4 generated version, Turnitin found that “32% of qualifying text in this submission has been determined to be generated by AI.” The detected text was highlighted and found to be the majority of section three of the assignment. Sections one and two were undetected. As a guideline, Turnitin’s current suggested threshold for faculty staff is that any submission that has a Turnitin AI detection greater than 20%, Turnitin are 99% confident that some of that submission has been generated by AI (Chechitelli, 2023; Nanda, 2023; Turnitin, 2023), with the usual caveat that ultimately the institution is responsible for checking and maintaining its own academic integrity.

Mapletoft N. et al. (2024). An attempt to cheat using GPT-4: findings, discussion and recommendations for academic staff and students. *Enhancing Teaching and Learning in Higher Education* Vol1. pp 52-73. <http://doi.org/10.62512/etlhe.11>

Findings and discussion of the GPT-4 essay by the marker

The assignment is a Framework for Higher Education (FHEQ) Level 5 (QAA, 2014, 2023c) essay for a module which is part of an undergraduate business and management degree. The learning outcomes and assignment tasks within it are consistent with Bloom's taxonomy (Bloom, 1956) and are assessed accordingly. The marking and grading criteria consist of: the overall academic standard of the submission, the impact of existing and emerging digital technologies on organisations, the analysis of knowledge management culture linked to the student's organisation, evidence of wider reading and research and appropriate Harvard referencing.

The AI-generated submission shows a high level of reading and research with almost every point qualified with a citation and reference. This would indicate that the assignment could achieve a good or very good mark overall. However, considering the domain of the task (digital technology) many of the references were potentially too old to allow this. Overall, the structure of the assignment was correct.

In terms of presentation the writing style was simplistic and very repetitive with no meaningful narrative for the reader to follow. It is highly descriptive with little in the way of critical analysis or evaluation to enable the reader to understand the relative merits or importance of different issues. For example, the essay says, "Here are some of the key impacts.." (OpenAI, 2023) without analysing or explaining which are more significant in this instance than others. There is no attempt to compare or contrast issues raised, which is very much what we would expect of an assignment such as this at this level. The essay also fails to explain key concepts such as knowledge management and important related ideas such as Communities of Practice, Knowledge Creating Company and Learning Organisation which is a significant oversight.

In terms of the task of providing recommendations, it treats the organisation in an entirely generic way which completely fails to satisfy the requirement of a contextual analysis of this sort. Overall, this submission would satisfy general FHEQ Level 4 (QAA, 2014, 2023c) requirements that demonstrate remembering and understanding very well but fails significantly in terms of higher-level skills such as critical analysis and evaluation.

Feedback to this student would have included that this appears to be a satisfactory first draft, but that much further work was required for it to achieve a better grade at this level.

The work rate in terms of reading and research is very impressive, but unfortunately it flatters to deceive. This assignment would achieve a pass at 40%-49%, largely through the contribution of the level of reading and research.

On a subjective note, after reading this assignment it would be relatively easy to identify similar artefacts as the style and structure of the writing is very distinctive and frankly robotic. Information is presented repetitively, with little emphasis or judgement and in a very bland style. It would certainly be a simple matter of comparing it to another sample of a student's work to judge whether they were likely to have produced it or not.

Alternatively, as a research exercise to produce a set of ideas that a student then uses to produce a more sophisticated and satisfying essay this could be a successful starting point. If the student applied their critical abilities to this material, refined and enhanced it in line with the assignment task this could be a successful approach. This would raise further issues in terms of the fact that it reduces the need for personal research activity and to some degree compromises established marking and grading criteria. Being able to assess the 'value added' that the student has achieved would become the new goal of the assessment. Perhaps one of the assessment tasks could be for the student to critically evaluate the output of the AI model before they began their development of it.

Further discussion

Turnitin say that one in ten essays that they checked during the first three months of their AI detector's operation, were found to contain substantial amounts of writing that was likely to have been created using artificial intelligence, (Williams, 2023b). One in thirty essays was flagged as having at least 80% AI generated content (Chechitelli, 2023). There is the need for institutions to be cautious of accepting the software's opinion on whether essays have been generated by AI however (Sokol, 2023), false accusations can be both mentally and financially damaging to the student. There is a difficulty for academic staff when they are presented with convincingly affirmed numbers by an AI detection tool, sometimes without asking for AI detection; they may wish to ignore the indication, knowing that it may be a false positive. However, it is hard to ignore it once it has been seen (Davis et al., 2023) due to the phenomenon of 'ironic process theory' or 'the white bear problem' based on Wegner's foundational work on thought suppression (Wegner et al., 1987).

AI detection tools have been shown to be 'biased against non-native English speakers', when seven widely used detection tools incorrectly labelled 61.3% of essays written by Chinese students as "AI-generated" (Williams, 2023a). There are also concerns that ChatGPT has been found to reproduce discriminatory associations across race, gender, disability and ethnicity (Bender et al., 2021; Lindebaum, 2023), due to the limitations of the data on which it has been trained.

Morozov (2014) called the idea that every problem has a solution, 'Solutionism'. However, technology is rarely the best way to solve our social problems (Shapiro, 2023), and our findings show that in practice, technology is not the best way to solve the issue of using AI to plagiarise; it is better to educate against plagiarism (Anders, 2023). Turin's proof by contradiction (Shapiro, 2023) shows us that technology does not work to solve our social problems in theory either.

Recommendations

This research has been undertaken to be used in educational practice, to support staff and students of UCQ developing our critical approach to the use and detection of generative AI tools. Transferability in this study relates to the likelihood that these research results are transferable to other HEPs. What follows is an overview of the key recommendations for academic staff and students.

Academic staff recommendations

In response to the increasing incorporation of AI technology in academia, our research provides several recommendations that academic staff can use to help them navigate through this transformative era. Similarly to Singh and Singh (2023) our study suggests that AI detection should not be assumed to give consistently accurate results, and so should not be relied upon exclusively.

It is recommended that academic staff stay up-to-date with industry developments, briefing papers and recommendations, for example the Quality Assurance Agency (QAA) guidance on ChatGPT and AI (QAA, 2023b).

Staff should include AI tools as part of the curriculum, including clear examples of what is and is not acceptable, linking with established academic integrity policy and digital competency (Anders, 2023).

Curriculum staff should look to optimise assignment question design using Bloom's taxonomy (Bloom, 1956) as a form of mitigation against a potential negative impact of ChatGPT on education (Elsayed, 2023).

Staff should consider alternative assessment methods by questioning the fundamental learning objectives for the ChatGPT era (QAA, 2023a). If AI can produce essays that are as good or even better than student work, perhaps essays are not the best method for assessment (Ross, 2023).

Staff can emphasise the importance of the person over the technology, so the individual takes the central role; a form of digital humanism (Fuchs, 2022; Werthner et al., 2022).

One way to test the reliability and validity of a student's knowledge is through a *viva voce* (Abuzied & Nabag, 2023). Just having this as an option may prevent a student from including information that does not come directly from their own thinking.

Where there is a case of infringement of academic integrity, warning letters have been shown to be effective as a way of correcting behavior for various misdemeanors (Brewer et al., 2019). Studies have however shown variance depending on the age and gender of the recipients (Jones, 1997), the method of message delivery (Moss & Kelcey, 2022) and whether the message is low-threat or high-threat, where sometimes the low-threat letter has been more effective (Shapiro, 2023).

Student recommendations

Students are encouraged to embrace AI as a useful tool and writing aid (Imran & Almusharraf, 2023) "AI will not replace you. A person using AI will" (Spanidis, 2023). There are an estimated three million papers across 30,000 journals published annually and AI-generated summaries of these papers can be a useful starting point to get an overview of papers (Williams & Grove, 2023).

Students should learn how to create detailed prompts in order to get the desired response from ChatGPT (Hunter, 2023).

Under no circumstances should students plagiarise. They must not assume that if they run an AI tool check, that this will give them accurate results or immunity from academic regress.

Students cannot assume that the AI-generated output is accurate, it is the student's responsibility to check. It cannot be assumed that any references provided are accurate, or that the referenced material even exists, ChatGPT has a tendency to "hallucinate" when asked to cite scholarly articles in a chosen discipline (Williams & Grove, 2023).

Students must not reference materials that they have not read. As a report author, the student is accountable for the work that they produce, this accountability cannot apply to an LLM (Nature, 2023).

In conclusion

This research reinforces the significance of SoTL in guiding educational practices in the age of AI-assisted writing. The article employs a SoTL approach by systematically investigating the use of ChatGPT in HE, reflecting on teaching practices, involving students and faculty in the research, and sharing the results to inform and improve teaching and learning. This revealed that, whilst AI can provide valuable insights and aid academic writing, and it can be very difficult to confidently detect, it cannot replace academic integrity of criticality.

In this era of technological advancement, SoTL serves as a reminder that education remains a human endeavor, where the exchange of ideas and the cultivation of critical thinking are paramount. SoTL reminds us to embrace AI as a tool for enhancement, not a shortcut, and to equip students with the skills to use AI responsibly. Ultimately, SoTL reinforces the central role of educators in guiding the ethical and meaningful integration of AI into the educational landscape.


References

Abuzied, A. I. H., & Nabag, W. O. M. (2023). Structured viva validity, reliability, and acceptability as an assessment tool in health professions education: A systematic review and meta-analysis. *BMC Medical Education*, 23(1), 531.

<https://doi.org/10.1186/s12909-023-04524-6>

Alexander, K., Savvidou, C., & Alexander, C. (2023). Who Wrote This Essay? Detecting Ai-Generated Writing in Second Language Education in Higher

Mapletoft N. et al. (2024). An attempt to cheat using GPT-4: findings, discussion and recommendations for academic staff and students. *Enhancing Teaching and Learning in Higher Education* Vol1. pp 52-73. <http://doi.org/10.62512/etlthe.11>

- Education. *Teaching English With Technology*, 2023(2).
<https://doi.org/10.56297/BUKA4060/XHLD5365>
- Ali, R., Tang, O. Y., Connolly, I. D., Zadnik Sullivan, P. L., Shin, J. H., Fridley, J. S., Asaad, W. F., Cielo, D., Oyelese, A. A., & Doberstein, C. E. (2023). Performance of ChatGPT and GPT-4 on neurosurgery written board examinations. *medRxiv*, 2023–03.
- Alshami, A., Elsayed, M., Ali, E., Eltoukhy, A. E. E., & Zayed, T. (2023). Harnessing the Power of ChatGPT for Automating Systematic Review Process: Methodology, Case Study, Limitations, and Future Directions. *Systems*, 11(7), 351.
<https://doi.org/10.3390/systems11070351>
- Anders, B. A. (2023). Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking? *Patterns*, 4(3), 100694. <https://doi.org/10.1016/j.patter.2023.100694>
- Barnett, R. (2022). *The philosophy of higher education: A critical introduction*. Routledge.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
<https://doi.org/10.1145/3442188.3445922>
- Bloom, B. S. (Ed.). (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals—Handbook I: Cognitive Domain*. McKay.
- Brewer, R., De Vel-Palumbo, M., Hutchings, A., Holt, T., Goldsmith, A., & Maimon, D. (2019). *Cybercrime Prevention: Theory and Applications*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-31069-1>
- Chechitelli, A. (2023, March 16). *Understanding false positives within our AI writing detection capabilities*. <https://www.turnitin.com/blog/understanding-false-positives-within-our-ai-writing-detection-capabilities>
- Coe, C. D. (2011). Scaffolded writing as a tool for critical thinking: Teaching beginning students how to write arguments. *Teaching Philosophy*, 34(1), 33–50.
- Corrector App. (2023). *AI Detector & Chat GPT Detector Free* [Computer software]. Corrector App. <https://corrector.app/ai-content-detector/>

- Curado, C., Oliveira, M., & Antunes, M. (2019). Organizational ambidexterity and customer relationship management: A cycle of virtue. *Knowledge and Process Management*, 26(3), 229–243. <https://doi.org/10.1002/kpm.1605>
- Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering*, 27(1), 113–118.
- Davis, J., Slade, C., Liu, B., & Zhegkui Wang, D. (2023, August 3). *Campus webinar: Artificial intelligence and academic integrity*. <https://www.timeshighereducation.com/campus/campus-webinar-artificial-intelligence-and-academic-integrity?>
- Duarte, F. (2023, July 13). *Number of ChatGPT Users (2023)*. Exploding Topics. <https://explodingtopics.com/blog/chatgpt-users>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Elsayed, S. (2023). *Towards Mitigating ChatGPT's Negative Impact on Education: Optimizing Question Design through Bloom's Taxonomy*. <https://doi.org/10.48550/ARXIV.2304.08176>
- Fuchs, C. (2022). *Digital humanism: A philosophy for 21st century digital society* (First edition). Emerald Publishing.
- GPTRadar. (2023). *Detect AI generated text in a click* [Computer software]. GPTRadar. <https://gptradar.com>
- GPTZero. (2023). *More than an AI detector Preserve what's human*. [Computer software]. GPTZero. <https://gptzero.me>
- Huawei, S., & Aryadoust, V. (2023). A systematic review of automated writing evaluation systems. *Education and Information Technologies*, 28(1), 771–795. <https://doi.org/10.1007/s10639-022-11200-7>

Hunter, N. (2023). *The art of prompt engineering with ChatGPT: A hands-on guide*.

Nathan Hunter.

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature.

Contemporary Educational Technology, 15(4), ep464.

<https://doi.org/10.30935/cedtech/13605>

Jacob, S. (2023). Best Free Plagiarism Checker. *Best Free Plagiarism Checker*.

[https://originality.ai/blog/best-free-plagiarism-](https://originality.ai/blog/best-free-plagiarism-checker#:~:text=Best%20Free%20Plagiarism%20Checker%20-%20Originality.AI)

[checker#:~:text=Best%20Free%20Plagiarism%20Checker%20–](https://originality.ai/blog/best-free-plagiarism-checker#:~:text=Best%20Free%20Plagiarism%20Checker%20-%20Originality.AI)

[%20Originality.AI](https://originality.ai/blog/best-free-plagiarism-checker#:~:text=Best%20Free%20Plagiarism%20Checker%20-%20Originality.AI)

Jennex, M. E., & Olfman, L. (2006). A Model of Knowledge Management Success:

International Journal of Knowledge Management, 2(3), 51–68.

<https://doi.org/10.4018/jkm.2006070104>

Jones, B. (1997). Age, gender and the effectiveness of high-threat letters: An analysis of Oregon's driver improvement advisory letters. *Accident Analysis & Prevention*,

29(2), 225–234. [https://doi.org/10.1016/S0001-4575\(96\)00076-0](https://doi.org/10.1016/S0001-4575(96)00076-0)

Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2023). Gpt-4 passes the bar exam. *Available at SSRN 4389233*.

Kim, M.-Y., & Wilkinson, I. A. G. (2019). What is dialogic teaching? Constructing, deconstructing, and reconstructing a pedagogy of classroom talk. *Learning, Culture and Social Interaction*, 21, 70–86.

<https://doi.org/10.1016/j.lcsi.2019.02.003>

Kirchner, J. H., Ahmad, L., Aaronson, S., & Leike, J. (2023, July 20). New AI classifier for indicating AI-written text We're launching a classifier trained to distinguish

between AI-written and human-written text. *New AI Classifier for Indicating AI-Written Text*. [https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-](https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text)

[text](https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text)

Lin, J. C., Younessi, D. N., Kurapati, S. S., Tang, O. Y., & Scott, I. U. (2023).

Comparison of GPT-3.5, GPT-4, and human user performance on a practice ophthalmology written examination. *Eye*, 1–2.

Lindebaum, D. (2023, May 2). *Researchers embracing ChatGPT are like turkeys voting for Christmas*. Times Higher Education (THE).

Mapletoft N. et al. (2024). An attempt to cheat using GPT-4: findings, discussion and recommendations for academic staff and students. *Enhancing Teaching and Learning in Higher Education* Vol1. pp 52-73. <http://doi.org/10.62512/etlthe.11>

- <https://www.timeshighereducation.com/blog/researchers-embracing-chatgpt-are-turkeys-voting-christmas>
- Massingham, P., & Al Holaibi, M. (2017). Embedding Knowledge Management into Business Processes. *Knowledge and Process Management*, 24(1), 53–71.
<https://doi.org/10.1002/kpm.1534>
- Morozov, E. (2014). *To save everything, click here: Technology, solutionism and the urge to fix problems that don't exist*. Penguin Books.
- Moss, B. G., & Kelcey, B. (2022). Words of Warning: A Randomized Study of the Impact of Assorted Warning Letters on Academic Probation Students. *Community College Review*, 50(3), 253–268.
<https://doi.org/10.1177/00915521221087285>
- Nanda, N. (2023, June 7). *Turnitin AI Guiding Principles*.
Nature. (2023). *Preparing your material* | *Nature Biotechnology*.
<https://www.nature.com/nbt/submission-guidelines/preparing-your-submission>
- OpenAI. (2023). *ChatGPT-4* [Computer software]. <https://chat.openai.com/chat>
- Originality.ai. (2023, July 25). *Accurate AI Content Detector and Plagiarism Checker*.
<https://originality.ai>
- QAA. (2023a). *Reconsidering assessment for the ChatGPT era: QAA advice on developing sustainable assessment strategies*.
- QAA. (2014, October). *The Frameworks for Higher Education Qualifications of UK Degree-Awarding Bodies*. <https://www.qaa.ac.uk/docs/qaa/quality-code/qualifications-frameworks.pdf>
- QAA. (2023b). *ChatGPT and Artificial Intelligence*.
<https://www.qaa.ac.uk/membership/membership-areas-of-work/academic-integrity/chatgpt-and-artificial-intelligence#>
- QAA. (2023c, May 10). *Quality Code*. The Quality Code. <https://www.qaa.ac.uk/the-quality-code>
- Ross, J. (2023, July 4). *ChatGPT forcing 'fruitful conversations' about assessment*. Times Higher Education (THE).
<https://www.timeshighereducation.com/news/chatgpt-forcing-fruitful-conversations-about-assessment>

- Shapiro, S. (2023). *Fancy bear goes phishing: The dark history of the information age, in five extraordinary hacks*. Allen Lane.
- Singh, A. (2023). A Comparison Study on AI Language Detector. *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, 0489–0493. <https://doi.org/10.1109/CCWC57344.2023.10099219>
- Singh, H., & Singh, A. (2023). ChatGPT: Systematic Review, Applications, and Agenda for Multidisciplinary Research. *Journal of Chinese Economic and Business Studies*, 21(2), 193–212. <https://doi.org/10.1080/14765284.2023.2210482>
- Sokol, D. (2023, July 10). *It is too easy to falsely accuse a student of using AI: A cautionary tale*. Times Higher Education (THE). <https://www.timeshighereducation.com/blog/it-too-easy-falsely-accuse-student-using-ai-cautionary-tale>
- Spanidis, G. (2023). *Write like a pro: The ChatGPT guide for crafting killer copywriting*. PublishDrive. <https://aicopymastery.com/>
- Taecharungroj, V. (2023). “What Can ChatGPT Do?” Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *Big Data and Cognitive Computing*, 7(1), 35. <https://doi.org/10.3390/bdcc7010035>
- Turnitin. (2023). *AI Writing Detection Frequently Asked Questions*. <https://www.turnitin.com/products/features/ai-writing-detection/faq>
- University of Cambridge. (2023). *Dialogic: Faculty of Education*. <https://www.educ.cam.ac.uk/research/programmes/dialogic/whatis.html>
- Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology*, 53(1), 5.
- Werthner, H., Prem, E., Lee, E. A., & Ghezzi, C. (Eds.). (2022). *Perspectives on Digital Humanism*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-86144-5>
- Williams, T. (2023a, July 11). *AI text detectors ‘biased against non-native English speakers’*. Times Higher Education (THE). <https://www.timeshighereducation.com/news/ai-text-detectors-biased-against-non-native-english-speakers>

Williams, T. (2023b, July 25). *Turnitin says one in 10 university essays are partly AI-written*. Times Higher Education (THE).

<https://www.timeshighereducation.com/news/turnitin-says-one-10-university-essays-are-partly-ai-written>

Williams, T., & Grove, J. (2023, May 15). *Five ways AI has already changed higher education*. Times Higher Education (THE).

<https://www.timeshighereducation.com/depth/five-ways-ai-has-already-changed-higher-education>

Willingham, D. T. (2007). Critical thinking: Why it is so hard to teach? *American Federation of Teachers Summer 2007*, p. 8-19.

Writer. (2023). *AI Content Detector* [Computer software]. Writer. <https://writer.com/ai-content-detector/>

ZeroGPT. (2023). *AI Text Detector* [Computer software]. ZeroGPT. <https://www.zerogpt.com>