

Can Artificial Intelligence Complete My Assessment? A Student Led Initiative to Stress Test the Academic Integrity of University Assessment Using Generative AI

Aidan Duane 

School of Business, South East Technological University (SETU), Waterford City, County Waterford, Ireland.

How to cite: Duane, A. 2024. Can Artificial Intelligence Complete My Assessment? A Student Led Initiative to Stress Test the Academic Integrity of University Course Work Using Generative AI. In: 10th International Conference on Higher Education Advances (HEAd'24). Valencia, 18-21 June 2024. <https://doi.org/10.4995/HEAd24.2024.17143>

Abstract

The ability of Generative AI (GenAI) to perform complex tasks has caused mixed feelings within the field of education. The most significant concern is the implications of GenAI for academic integrity. In this study, students applied GenAI to complete past university assessments adapted as research tests, with the goal of achieving a pass grade when graded by an academic, undetected by an AI writing detection tool. The study reveals that from a sample of 21 valid research tests, 23.8% (5) passed when graded by an academic achieving grades between 40% and 60%, with AI writing detection scores ranging between 0% and 14%. AI writing detection scores lower than 20% are potentially a false positive in the context of investigating breaches of academic integrity. The researcher concludes that many traditional methods of assessment in universities are obsolete in the face of increasingly undetectable AI generated solutions.

Keywords: *Generative Artificial Intelligence; GenAI; Assessment; Academic Integrity; Artificial Intelligence Writing Detection.*

1. Introduction

Generative AI (GenAI) known as ChatGPT developed by OpenAI, and released publicly in November 2022, has become the fastest growing consumer application in history (Marr, 2023). The ability of GenAI to perform complex tasks (Liu et al., 2023; McKinsey, 2023) within the field of education has caused mixed feelings among educators (García-Peñalvo, 2023; Halaweh, 2023; Kohnke et al., 2023; Malinka et al., 2023; Rudolph and Tan, 2023; Sawahel, 2023). Integrating GenAI in education raises concerns about assessment and evaluation, as some traditional methods may become obsolete in the face of GenAI solutions (Fowler, 2023; García-Peñalvo, 2023; Hockly, 2023; Lancaster, 2021). The most significant concern about GenAI in

the field of education, is the implications of GenAI for academic integrity (Firat, 2023b; Fowler, 2023; Lancaster, 2021; Malinka et al., 2023; Roe and Perkins, 2022). This paper presents the results of a study where students applied GenAI to complete past university assessments, adapted as research tests, with the goal of achieving a pass grade when graded by an academic, undetected by AI writing detection tools.

2. Literature Review

Artificial Intelligence (AI) is now firmly rooted in our daily lives. Since November 2022 a form of generative AI (GenAI) known as ChatGPT developed by OpenAI, has become the fastest growing consumer application in history. A report by Ofcom United Kingdom (Ofcom, 2023) reveals that 80% of British teenagers aged from 13-17 years old, and 40% of children between the ages of 7-12 years old have used GenAI tools and services for schoolwork or leisure. ChatGPT is the most widely used GenAI tool among internet users aged 16 and over (Ofcom, 2023). Education features strongly in the target audience interests of ChatGPT site visitors (Similarweb, 2023). ChatGPT's journey from concept to influential GenAI model exemplifies this rapid evolution of GenAI. This groundbreaking model has driven progress in GenAI development and spurred transformation of work practices across a wide range of industries (Marr, 2023). Thus, given the potential impact of GenAI on business and indeed all facets of life, it is critical that this technology becomes part of the educational experience of students particularly as it heavily influences their future careers.

The ability of GenAI to perform complex tasks within the field of education has caused mixed feelings among educators as it disrupts existing educational praxis (Baidoo-Anu and Owusu Ansah, 2023). However, as studies are only emerging with respect to the uses and benefits of GenAI for education, it is not yet possible to assert a consensus among academics with specific reference to GenAI in education (Firat, 2023a). Integrating GenAI in education raises concerns about assessment and evaluation, as traditional methods may become obsolete in the face of GenAI solutions (Rudolph and Tan, 2023; Sullivan et al., 2023). Furthermore, as GenAI can contain factual inaccuracies and biases (Baidoo-Anu and Owusu Ansah, 2023; Sullivan et al., 2023), it can only enhance student learning if academics are experts in their field and can detect such inaccuracies and biases. Furthermore, to use GenAI as a tool, educators must address significant gaps in digital competencies to use it ethically and effectively to support learning (Kohnke et al., 2023). GenAI education literature also raises significant concerns with respect to academic integrity (Amini-Salehi, 2023; An et al., 2023; Malinka et al., 2023; Zhai, 2023).

While OpenAI released an AI writing detection tool in February 2023, it was withdrawn as it was ineffective in July 2023 (OpenAI, 2023a). In April 2023, Turnitin launched an AI writing detection tool embedded in the Turnitin Feedback Studio (Caren, 2023). The Turnitin AI writing detection tool quickly established itself as a key tool for academics to detect GenAI writing in

assessments including essays, presentation slides, and narrative analyses. However, this AI writing detection tool is fallible and can generate false positives (Fowler, 2023).

The ability of GenAI to paraphrase texts and reduce plagiarism detection (Amini-Salehi, 2023) also raises concerns as it conceals work that is not original and results in a distorted perception of the student's academic writing abilities. Wang (2023) found that college students with higher metacognitive levels were better able to describe their goals and processes using GenAI prompt engineering and were better able to critically adapt the answers given by GenAI. By contrast, students with lower metacognitive levels were found to rely more heavily on GenAI, rather than just using it as a support tool. Thus, GenAI can potentially provide some students with an advantage over others, particularly if it is used to conceal plagiarism, or indeed if it is used surreptitiously in the generation of an entire solution to an assessment. GenAI also raises concerns that students may outsource assessment tasks to those capable of rapidly producing high quality outputs (Zhai, 2023). Despite this important debate, extant literature pertaining to GenAI use among students is scarce in the field of education. Furthermore, the student's voice is poorly represented in research to date (Sullivan et al., 2023).

3. Research Method

This study explores the existing capabilities of students to use GenAI to complete academic assessment. The implementation of the study is quite simple and easily replicated. The Principal Investigator (PI) first invited university academics to submit a sample of past assessments from 2020-2022. The past assessments were from modules including Fund Reporting and Risk Management; Personal, Professional and Academic Skills; Business Research and Communication Skills; Business Strategy; Global Business Ethics; Organisational Behaviour; Management Skills; Professional Development; and Behavioural Finance. The assessments included written essays; analysis of spreadsheet data; written reports; presentation slides; reflective learning diaries; and discursive analysis. These past assessments were then adapted as research tests removing cover sheets and submission details/dates but preserving the original questions/instructions. Twenty-six (26) research tests were generated from the past assessments.

The PI posted invitations on the university course management systems (CMS) for students to voluntarily enroll in two GenAI research events in April and September 2023. The goal of these GenAI research events was to challenge students to apply GenAI to complete the past assessments adapted as research tests, with the goal of achieving a pass grade when graded by an academic, undetected by GenAI writing detection tools. Each research event had a duration of 2 hours. This is important as the original assessments gave students several weeks to complete their submissions. The students were all 1st year, 2nd year and 3rd year business students.

The students submitted their research test solutions to the CMS. Two (2) of the twenty-six (26) research tests submitted were spoiled and inadmissible in the analysis. The valid research tests (24) were processed using GenAI writing detection software to determine an AI writing detection score. The PI then distributed the research tests for grading using the original marking scheme to the academics who volunteered the original past assessment samples. The academics were not provided with the GenAI writing detection score when grading the tests. Of the 24 valid research tests, 3 were not graded by the academic who provided the original assessment due to time constraints. Thus, the results of the data analysis are confined to 21 valid research tests from the two GenAI research events. Once the 21 research tests were graded by the academics, the PI held informal interviews lasting 40 minutes to 1 hour, with each academic to discuss their experiences of the research tests, the grades they awarded, and their thoughts on the challenges of GenAI for academic integrity.

The PI considers this research methodology an appropriate and easily replicated benchmarking test of GenAI and GenAI writing detection tools, so that as academics, we can establish a clear understanding of the challenges GenAI poses for academic integrity.

4. Results

The AI writing detection tool vendor warns that low AI Detection Scores (less than 20% GenAI writing detection) have a higher likelihood of being false positives and are thus insufficient grounds upon which to raise an academic integrity enquiry. In this study, the AI writing detection tool was successful in detecting high AI Detection Scores (>20%) in 75% (18/24) of the research tests. Thus, all of these tests would have been flagged to the academic and would have warranted further investigation for breaches of academic integrity.

The analysis shows that 25% (6/24) of the research tests achieved an AI Detection Score of 14% or less, and would not have been flagged to the academic, and would not have been investigated for breaches of academic integrity. Analysis of these six (6) research tests, reveals that four (4) achieved a 0% AI Detection Score, while two research tests yielded 10% and 14% AI Detection Scores. Three (3) of these research tests were submitted by first year students and three (3) were submitted by second year students. None of the third year students submitted research tests that evaded the AI writing detection tool.

Five (5) of the six (6) research tests with no/low AI Detection Scores received a pass grade from the academic who provided the past assessment. This equates to 23.8% of all test submissions corrected by an academic (n=21). The average grade for research tests achieving a no/low AI Detection Score and achieving a pass grade from the academic was 53%, with 60% the highest grade and 40% the lowest grade. It is worth noting that one other research test submitted by a first year student achieved a 0% detection rate but only achieved a 32% grade from the

academic. Thus, although the research test received a fail grade, the GenAI content was not detected by the AI writing detection tool. It is important to highlight that the five (5) (23.8%) research tests with no/low AI Detection Scores that achieved a pass grade were not attributable to a specific academic, module or submission type as they were evenly dispersed.

The AI writing detection rate decreased significantly from 90% to just 37% between GenAI Research Event 1 in April 2023 and GenAI Research Event 2 in September 2023. It is also critical to highlight that the five (5) research tests (23.8%) that achieved no/low AI Detection Scores were submitted during AI Research Event 2. This may be an indication of how well GenAI has been adopted and mastered by students since its launch in November 2022, and the PI holding AI Research Event 1 in April 2023 and AI Research Event 2 in September 2023.

In this study, GenAI struggled to work with Microsoft Excel spreadsheet analysis. None of the research tests passed when graded by the academic. Initially trained on data up to September 2021, ChatGPT's training set continues to be upgraded and it is now integrated with Microsoft Bing search engine. Initially, ChatGPT responded to just text based questions, but it now incorporates image-based requests, and its integration with Microsoft Office and Microsoft Bing now enables Microsoft Excel Spreadsheets to incorporate micro-app AI agents capable of extracting a table of data, cleaning up the data, and analysing the data (OpenAI, 2023b). Thus, the difficulties encountered by students undertaking research tests requiring analysis of Microsoft Excel data in this study no longer exist. Thus, GenAI poses an ever-greater challenge to academic integrity if these issues are not considered when assessments are being set.

Following the collection and analysis of all AI Detection Scores and research test grades, the PI held informal meetings with each participant academic. These academics contend that as GenAI improves, and students increasingly engage with it, academics will have to be subject matter experts to detect GenAI content. However, a counter-argument could be that with the evolution of GenAI, will it indeed replace the need for any academic to be a subject matter expert? Participant academics were concerned that students can use GenAI to create a bibliography to accompany GenAI text, and that some or indeed all the bibliography may not be authentic. Participant academics are concerned they do not have the resources or time to follow every reference to verify if it is authentic, or indeed relevant to the text. Participant academics believe that fundamental questions need to be asked about what we are teaching, why we are teaching it, how we are teaching it, and what careers we are preparing graduates to enter. Students commencing university in 2023-2024 will emerge from degree programmes into an employment landscape transformed by AI, and much of what they are learning could be redundant if roles are supplanted by AI. Thus, participant academics contend that a bigger conversation must occur about the impact of AI on education beyond that of its impact on academic integrity.

5. Limitations

While all the aims, objectives and tasks were completed as per the research funding proposal, the study encountered several limitations which impacted its potential output including low levels of academic staff engagement and a lack support for the study by some senior academic managers. Fear of AI is pervasive, and some academics and academic managers appear to have adopted a head in the sand approach to the impact of AI on their teaching and learning strategies and indeed the impact of AI on the entire education system.

6. Conclusion

GenAI provides educators with an opportunity to redirect the achievement of learning outcomes through students engaging in written tasks, to instead focusing on how students develop higher-order critical thinking skills and mentoring their learning and progress (Sullivan et al., 2023). Educators must explore GenAI applications to significantly improve inclusion for people with communication disabilities (Hemsley et al., 2023; Starcevic, 2023; Sullivan et al., 2023) and enhance participation and success for students from disadvantaged backgrounds (Sullivan et al., 2023). However, educators must first address the significant gaps in digital competencies for ethical and effective use of GenAI to support learning (Kohnke et al., 2023).

From a strategic educational perspective, GenAI compels educators to adapt teaching, learning and assessment practices that incorporate the new reality of living, working, and studying in a world where GenAI is widely available (Halaweh, 2023; García-Peñalvo, 2023; Liu et al., 2023; Rudolph and Tan, 2023; Sullivan et al., 2023). It is essential that policy makers, researchers, educators, and technology experts work together and start conversations on how these evolving GenAI tools can be used safely and constructively to improve education, and support learning while maintaining academic integrity. In UK and US universities, working groups have already been set up to assess the challenge of GenAI. Provisional outcomes indicate that methods of teaching, learning and assessment will have to radically change. This study provides tangible evidence that GenAI is disrupting education, and the global education system must adapt to meet these disruptive challenges and opportunities.

Acknowledgements

The PI would like to thank the senior management, academics and students who supported and participated in this study, without whose engagement and teamwork, it could not have been completed. The PI and the participants would like to acknowledge the funding provided by the Strategic Alignment of Teaching and Learning Enhancement Funding in Higher Education 2022/2023 and the N-TUTORR National Project 2023 without whose support this research would not have been possible.

References

- Amini-Salehi, E., Hassanipour, S., Bozorgi, A., Keivanlou, M.H., Dave, T., Alotaibi, A., Mellatdoust, P., Joukar, F., & Bakhshi, A. (2023). The ability of ChatGPT to paraphrase texts and reduce plagiarism. *JMIR Medical Education*. JMIR Preprints #53308: <https://preprints.jmir.org/preprint/53308>.
- An, Yingyi, Ouyang, W., & Zhu, F. (2023). ChatGPT in higher education: Design teaching model involving ChatGPT. *Lecture Notes in Education Psychology and Public Media*, 24, 47-56. <http://dx.doi.org/>.
- Baidoo-Anu, D. & Owusu Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning, January 25. <http://dx.doi.org/10.2139/ssrn.4337484>.
- Caren, C. (2023). The launch of Turnitin's AI writing detector and the road ahead, 4th April. *Turnitin*. <https://www.Turnitin.com/blog/the-launch-of-Turnitins-ai-writing-detector-and-the-road-ahead>.
- Firat, M. (2023a). What ChatGPT means for universities: Perceptions of scholars and students, *Journal of Applied Learning & Teaching*, 6(1). <https://doi.org/10.37074/jalt.2023.6.1.32>
- Firat, M. (2023b). How ChatGPT can transform autodidactic experiences and open education? OSFPreprints, <https://doi.org/10.31219/osf.io/9ge8m>
- Fowler, G. (2023). We tested a new ChatGPT-detector for teachers. It flagged an innocent student, April 3, *The Washington Post*. <https://www.washingtonpost.com/technology/2023/04/01/chatgpt-cheating-detection-Turnitin/>.
- García-Peñalvo, F.J. (2023). The perception of artificial intelligence in educational contexts after the launch of ChatGPT: Disruption or panic? *Education in the Knowledge Society*, 24, 1-9. <https://doi.org/10.14201/eks.31279>.
- Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, 15(2). <https://doi.org/10.30935/cedtech/13036>.
- Hockly N (2023). Artificial intelligence in English language teaching: The good, the bad and the ugly. *RELC Journal* 54(2): 445–451. <https://doi.org/10.1177/00336882231168504>.
- Kohnke, L., Moorhouse, B.L., & Zou, D. (2023). ChatGPT for language teaching and learning, *RELC Journal*. <https://doi.org/10.1177/00336882231162868>.
- Lancaster, T. (2021). Academic dishonesty or academic integrity? Using Natural Language Processing techniques to investigate positive integrity in academic integrity research. *Journal of Academic Ethics*, <https://doi.org/10.1007/s10805-021-09422-4>.
- Liu, A., Bridgeman, D., & Miller, B. (2023). As uni goes back, here's how teachers and students can use ChatGPT to save time and improve learning. *The Conversation*, February 28. <https://theconversation.com/as-uni-goes-back-heres-how-teachers-and-students-can-use-chatgpt-to-save-time-and-improve-learning-199884>.
- Malinka, K., Peresíni, M., Firc, A., Ondrej, H., & Janus, F., (2023). On the educational impact of ChatGPT: Is AI ready to obtain a university degree? In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education*. Association for Computing Machinery, NY, 47-53. <https://doi.org/10.1145/3587102.3588827>.

- Marr, B. (2023). A short history of ChatGPT: How we got to where we are today. *Forbes*, May 19. <https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/>.
- McKinsey (2023). The State of AI in 2023. *McKinsey Group*. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year>.
- Ofcom (2023). Gen Z driving early adoption of Gen AI. *Ofcom*. <https://www.ofcom.org.uk/news-centre/2023/gen-z-driving-early-adoption-of-gen-ai>.
- OpenAI (2023a). Due to high demand, we've temporarily paused upgrades. *OpenAI Developer Forum*, Nov 21st. <https://community.openai.com/t/due-to-high-demand-weve-temporarily-paused-upgrades-when-will-chatgpt-4-be-available-for-purchasing-again/517248?page=4>.
- OpenAI (2023b). Chat GPT for Excel. *OpenAI Developer Forum*, <https://community.openai.com/t/chat-gpt-for-excel-use-gpt3-inside-excel-sheets/86697/13>.
- Roe, J. & Perkins, M. (2022). What are automated paraphrasing tools and how do we address them? A review of a growing threat to academic integrity. *International Journal for Educational Integrity*, 18(1), pp.1-10. <https://doi.org/10.1007/s40979-022-00109-w>.
- Rudolph, J. & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 342-363. <https://doi.org/10.37074/jalt.2023.6.1.9>.
- Sawahel, W. (2023). Embrace it or reject it? Academics disagree about ChatGPT. *University World News*, <https://www.universityworldnews.com/post.php?story=20230207160059558>.
- Similarweb (2023). ChatGPT Demographics. *Similarweb*. <https://www.similarweb.com/website/chatgpt.com/#overview>.
- Sullivan, M., Kelly, A. & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching*, Vol. 6, No.1, March 20th. <https://doi.org/10.37074/jalt.2023.6.1.17>.
- Wang, L (2023). Metacognitive strategies used in ChatGPT among college students. *Lecture Notes in Education Psychology and Public Media*, <https://doi.org/10.54254/2753-7048/21/20230170>.
- Zhai, X. (2022). ChatGPT user experience: Implications for education. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4312418>.