

# Student Learning Performance Evaluation: Mitigating the Challenges of Generative AI Chatbot Misuse in Student Assessments

Chun Meng Tang<sup>1</sup> and Lee Yen Chaw<sup>2</sup>

<sup>1</sup>James Cook University Singapore, Singapore

<sup>2</sup>UCSI University, Malaysia

[chunmeng.tang@jcu.edu.au](mailto:chunmeng.tang@jcu.edu.au)

[chawly@ucsiuniversity.edu.my](mailto:chawly@ucsiuniversity.edu.my)

**Abstract:** Since the launch of ChatGPT, a growing number of generative artificial intelligence (AI) chatbots have entered the market. Although chatbots have the potential to help students learn, misusing them to complete assessments raises questions about the authenticity of the work and puts students at risk of academic misconduct. Given the crucial role of assessments in evaluating students' learning performance, uncertainties about the authenticity of the work call into question the extent to which students have achieved the intended learning outcomes. This study conducted a thematic analysis to provide an overview of the challenges that chatbot misuse may pose to student learning performance evaluation, followed by the various mitigation strategies to overcome these challenges. This study searched the Education Resources Information Centre (ERIC) database for peer-reviewed articles published in scholarly journals after 30 November 2022 (the launch date of ChatGPT), as this study focuses on generative AI rather than other types of AI) and until 30 April 2024. The thematic analysis of 17 articles identified five major themes (and respective sub-themes) in the discussions of these articles, i.e., reasons students use chatbots for assessments, challenges that chatbots may pose to student learning performance evaluation, mitigation strategies, detection strategies, and counter-detection strategies. As chatbots become more prevalent and powerful, the study's findings provide education stakeholders with insightful information on the implications of students misusing chatbots for assessments and how this affects their learning performance evaluation.

**Keywords:** Assessments, Education, Education policy, Generative AI, Learning erformance, Performance evaluation

---

## 1. Introduction

Generative artificial intelligence (AI) chatbots' (e.g., Google's Gemini, Microsoft's Copilot, and OpenAI's ChatGPT) ability to easily generate content for a wide range of topics at the command of a prompt has made such chatbots a popular tool among students to help with their assessments (Ansari et al., 2023; Ayanwale and Ndlovu, 2024; Tang and Chaw, 2023). Students can benefit from leveraging chatbots for tasks such as paraphrasing text, generating ideas, searching for information, etc. (Farrokhnia et al., 2024; Nugroho et al., 2024). Chatbots can quickly complete these tasks, despite concerns about the accuracy and originality of their outputs (Altun et al., 2024; Khalil and Er, 2023).

However, some students may misuse chatbots to complete their assessments (Crompton and Burke, 2024; Kooli, 2023), e.g., by copying and submitting chatbot output as their original work. Because assessments typically measure students' levels of achievement in the intended learning outcomes (Gao et al., 2020; Gerritsen-van Leeuwenkamp et al., 2019), misusing chatbots to complete assessments not only undermines students' ability to learn essential skills but also distorts assessment grades or marks, misrepresenting the students' true learning performance (Kooli, 2023).

Although academic literature has discussed the impacts of chatbots on student assessment, these discussions often occur in silos. Therefore, this study intends to perform a thematic analysis to synthesise the discussions, identify themes, and explore the relationships between these themes. Specifically, this study aims to provide a deeper and expanded answer to two research questions: (1) What are the challenges chatbots may pose to student learning performance evaluation if students misuse chatbots to complete their assessments, and (2) What are the measures to mitigate these challenges?

The study's findings contribute to the overall understanding of the implications of misusing chatbots for assessments and provide stakeholders in the education system, such as educators, students, policymakers, and administrators, with an overview of the challenges that misusing chatbots for assessments may pose to student learning performance evaluation, as well as mitigation strategies to address these challenges.

The remainder of this paper provides a background to the research, explains the research design and method, presents the study's findings, and concludes with a discussion of the study's findings and future research directions.

## **2. Research Background**

Student assessments are an essential part of the teaching and learning process. Depending on the purpose and objective of the assessment, educators typically use a variety of assessment methods, including reports, presentations, projects, exams, case studies, tests, and so on, to evaluate student learning performance to determine quality of work, progress, and areas for improvement (Flores et al., 2015; Neto et al., 2023). Broadly, assessment methods can be categorised into two types: formative and summative. Formative assessments provide information about students' progress towards the intended learning outcomes, whereas summative assessments measure student achievement of the intended learning outcomes (Lau, 2016).

Accurate evaluation of student learning performance provides meaningful feedback for continuous improvement in teaching and learning, policymaking, and governance (Al-Thani et al., 2014; Chaudhry et al., 2023). Educators and education institutions use the results of assessments to evaluate the effectiveness and quality of teaching and learning. They then implement necessary changes, such as modifying teaching and learning policies, strategies, activities, and assessments, to improve future student performance in achieving the learning outcomes (Biggs, 2014; O'Neill and Padden, 2022).

To accurately reflect the student's learning performance, it is essential that the assessment is the student's own work and that the student has put in their best effort to complete it. However, the emergence of chatbots, which can generate answers to nearly any question despite concerns about their accuracy, exposes educators and education institutions to the possibility of students misusing these chatbots and submitting the generated answers as their own (Dwivedi et al., 2023). This misuse of chatbots for assessments not only raises concerns about academic dishonesty, but also challenges the fundamental purpose and objective of student assessment, potentially misleading educators and education institutions about students' true learning performance (Kooli, 2023).

## **3. Research Design and Method**

### **3.1 Thematic Analysis**

Thematic analysis is a qualitative data analysis method that aims to identify and interpret repetitive patterns (themes) in the data, and the relationships between those themes, that are underlying certain aspects of a phenomenon (Boyatzis, 1998; Braun and Clarke, 2006). This study followed the reflexive thematic analysis approach as outlined in Braun and Clarke's (2006) six-phase guide: (1) familiarising with data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining themes, and (6) producing the report.

### **3.2 Database and Article Screening Process**

This study searched for articles that expressly addressed the challenges that generative AI may pose to student assessment in an educational setting. To find these articles, this study searched the Education Resources Information Centre (ERIC) database via ProQuest. This study did not limit the search to any specific educational disciplines, levels, or assessment methods, as the intention was to find articles discussing assessments in the broader context of chatbots. The search utilised a specific search string (see Figure 1 for the search string) and restricted the results to peer-reviewed, scholarly journals. This study conducted the search on 01 May 2024, and as the focus was on generative AI, limited the search to articles published after 30 November 2022 (ChatGPT's initial launch date). The search returned 120 articles.

To screen articles that were eligible for the thematic analysis, this study first examined the titles and abstracts of individual articles, and 92 articles were rejected based on several specific criteria (see Figure 1 for the criteria). This study then read the full text of the remaining 28 articles. This process removed another 12 articles. In addition, the analysis identified three more articles via citations in some of the articles. This study reviewed these additional articles but eliminated two of them. In total, this study deemed 17 articles eligible and included them in the thematic analysis. Figure 1 depicts the flow diagram of the screening process.

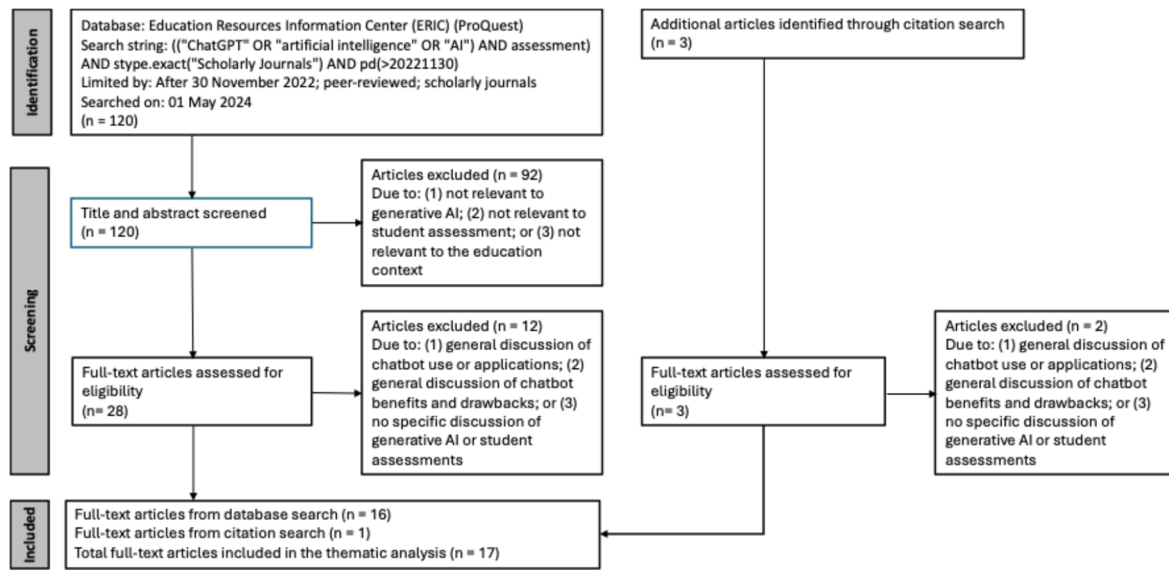


Figure 1: Flow diagram of the screening process (adapted from the PRISMA 2020 statement (Page et al., 2021))

### 3.3 Data Analysis

The dataset consisted of the articles identified through the screening process. Each article is deemed a data item in thematic analysis. After reading the individual articles to have a broad understanding of what they discussed, this study began extracting data that described certain aspects that were relevant to the research questions. The data extracts were typically a partial or complete sentence or paragraph. This study then coded the individual data extracts. It is possible to code a data extract one or more times to different codes. To maintain coding consistency, the coder was the same person who diligently applied the codes.

This study generated the codes following an inductive (data-driven) approach to look for thematic similarities across the data items (Braun and Clarke, 2006; Braun and Clarke, 2021). The inductive approach works without referring to an a priori framework but to look for themes that reflect the data. This study generated a preliminary list of 34 codes during data analysis.

A subsequent review of the codes indicated that some of them shared similar characteristics with an underlying theme. As a result, this study followed an iterative approach of merging these similar codes, reducing the number of codes to 19, and then categorising the 19 codes into five main themes. Each main theme consisted of several sub-themes, respectively. To validate the codes (which are also considered sub-themes), this study reviewed them again, both at the level of the coded data extracts and the dataset. It appeared that these codes reflected coherently the underlying patterns in the data.

## 4. Study Findings

This study identified five main themes in the data, i.e., reasons for using chatbots for assessments, challenges that chatbots may bring to student learning performance evaluation, mitigation strategies, detection strategies, and counter-detection strategies. There appears to be a relationship between the themes. Figure 2 depicts the thematic map, which includes the final five main themes and their sub-themes (denoted by bulleted points). There are 19 sub-themes.

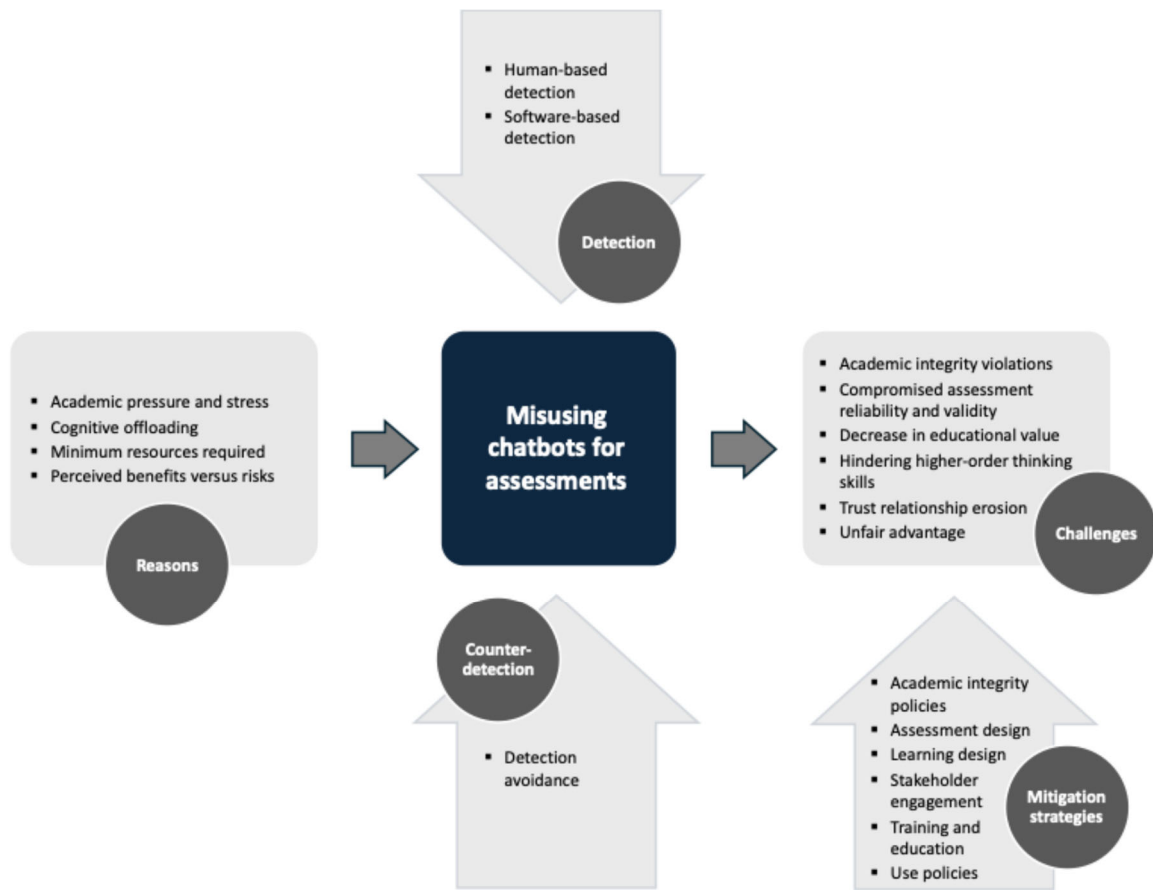


Figure 2: Thematic map (bulleted points denote sub-themes)

#### 4.1 Reasons for Using Chatbots for Assessments

Some discussions in the articles indicated four main reasons why students leveraged chatbots for help with their assessments: (1) cognitive offloading (e.g., Alexander et al., 2023; Bai et al., 2024; Perkins, 2023); (2) academic pressure and stress (e.g., Crawford et al., 2023; Perkins et al., 2024; Yeadon et al., 2023); perceived benefits versus risks (e.g., Adilov et al., 2024; Nikolic et al., 2023); and minimum resources required (e.g., Bai et al., 2024; Lancaster, 2023; Yeadon et al., 2023).

#### 4.2 Challenges That Chatbots may Pose to Student Learning Performance Evaluation

Most of the discussions in the articles focused on academic integrity violations (e.g., Lancaster, 2023; Nikolic et al., 2023; Perkins et al., 2024). Some specific examples include misinformation (Alexander et al., 2023; Chaudhry et al., 2023); questionable references (Bai et al., 2024; Niloy et al., 2024); plagiarism (Elkhatat, 2023; Ibrahim, 2023); biases (Alexander et al., 2023; Mao et al., 2024); cheating (Adilov et al., 2024; Chaudhry et al., 2023); misuse (Crawford et al., 2023; Mao et al., 2024); academic misconduct (Lancaster, 2023; Perkins, 2023); originality (Elkhatat, 2023); and authentication (Crawford et al., 2023).

A large percentage of discussions highlighted the possibility that chatbots, if misused, could hinder students from learning higher-order thinking skills such as critical thinking, creativity, analytic thinking, problem solving, and knowledge application (e.g., Bai et al., 2024; Blanke, 2024; Chaudhry et al., 2023; Mao et al., 2024; Niloy et al., 2024). Several discussions underlined that students could use chatbots as a shortcut tool (Blanke, 2024), outsourcing cognitive skills to chatbots (Chaudhry et al., 2023) and thus failing to satisfy learning outcomes (Lancaster, 2023). Some discussions, on the other hand, emphasised that if students did not gain essential skills from their learning (Chaudhry et al., 2023) or did not achieve the learning objectives (Alexander et al., 2023), or if the reputation of education institutions was in doubt (Crawford et al., 2023), the perceived value of education could decline.

Some discussions focused on compromised assessment reliability and validity when it was uncertain whether the assessments truly reflected student learning performance or achievement (e.g., Chaudhry et al., 2023; Gorichanaz, 2023; Ibrahim, 2023). A few discussions also raised concerns that students could gain an unfair

advantage by misusing chatbots to complete their assessments, even if unintentionally, and perform better than their peers who did not (e.g., Chaudhry et al., 2023; Perkins et al., 2024; Yeadon et al., 2023).

Other discussions stated that uncertainty about the authenticity of students' work (Blanke, 2024) could have potential adverse effects on trust relationships between students and lecturers, as well as between humans and technology (e.g., Bai et al., 2024; Mao et al., 2024; Gorichanaz, 2023) and could lead to accusations of academic misconduct (e.g., Gorichanaz, 2023; Perkins, 2023; Lancaster, 2023).

#### **4.3 Mitigation Strategies**

Many of the discussions in the articles suggested designing assessments to make it more difficult for students to misuse chatbots. Some specific examples include oral or physical assessment (Blanke, 2024; Gorichanaz, 2023; Yeadon et al., 2023); authentic assessment (Gorichanaz, 2023; Kortemeyer, 2023; Nikolic et al., 2023); skill-based assessment (Nikolic et al., 2023); critique of chatbot responses (Blanke, 2024; Kortemeyer, 2023; Perkins et al., 2024); interview-based assessment (Nikolic et al., 2023); specific requirements (e.g., discipline, framework, pre-approved topics) (Lancaster, 2023; Perkins et al., 2024); and podcasts (Crawford et al., 2023). Some discussions also called for incorporating AI as part of the assessment (e.g., Lancaster, 2023; Perkins et al., 2024) and developing assessment policies (including rubrics) (Alexander et al., 2023).

Some discussions emphasised the importance of institutional academic integrity policies (e.g., Alexander et al., 2023; Chaudhry et al., 2023; Lancaster, 2023), honour codes (e.g., Elkhataf, 2023), and the development of student moral characters (e.g., Crawford et al., 2023). Other discussions also highlighted the need for use policies or guidelines that promote the ethical, safe, and transparent use of AI (e.g., Gorichanaz, 2023; Mao et al., 2024; Perkins et al., 2024).

Concerning academic integrity and use policies, many of the discussions in the articles also stressed the need for educators and students to receive training and education on the ethical use of chatbots and their limitations (e.g., Gorichanaz, 2023; Lancaster, 2023; Mao et al., 2024), academic integrity policies (e.g., Chaudhry et al., 2023), and detection of AI-generated text (e.g., Alexander et al., 2023; Perkins et al., 2024).

Furthermore, some discussions pointed out the importance of designing learning to incorporate AI to teach students the AI skills they would need in the future (e.g., Bai et al., 2024; Kortemeyer, 2023; Nikolic et al., 2023). Such new skills could include prompt engineering (Alexander et al., 2023; Crawford et al., 2023), AI literacy (Mao et al., 2024), ethics (Nikolic et al., 2023), and greater emphasis on analysis and critical thinking (Blanke, 2024).

Other discussions touched on the need to engage stakeholders to better understand their perspectives on the use of AI. These stakeholders should include educators, students, education institutions, AI developers, etc. (e.g., Bai et al., 2024; Chaudhry et al., 2023; Lancaster, 2023).

#### **4.4 Detection Strategies**

Most of the discussions in the articles outlined the use of software to detect misuse of chatbots but raised concerns regarding its reliability and accuracy (e.g., Adilov et al., 2024; Ibrahim, 2023; Perkins et al., 2024). While software tools could detect AI-generated text, detecting AI-generated ideas proved more difficult (Gorichanaz, 2023). Some discussions focused on using human judgement to detect misuse (e.g., Chaudhry et al., 2023; Lancaster, 2023; Mao et al., 2024) and training educators to recognise AI-generated text by identifying specific features, such as questionable references or complex language (Perkins et al., 2024), spelling mistakes, or sentence construction (Alexander et al., 2023).

#### **4.5 Counter-Detection Strategies**

Some discussions in the articles noted how some students attempted to evade detection of their chatbot misuse by employing various strategies (e.g., Gorichanaz, 2023; Nikolic et al., 2023), such as prompt engineering or writing short sentences in bullet style (Perkins et al., 2024).

### **5. Discussion and Conclusions**

The study's findings indicate that because chatbots are user-friendly and cost-effective, students are inclined to use them to reduce their cognitive load (Alexander et al., 2023; Bai et al., 2024; Perkins et al., 2024). Consequently, students who are experiencing academic pressure and stress may be more inclined to misuse chatbots (Crawford et al., 2023; Perkins et al., 2024; Yeadon et al., 2023). These findings are consistent with Alshurafat et al.'s (2024) findings that factors such as pressure and technology accessibility may lead to chatbot misuse. Besides, students who misuse chatbots may be willing to prioritise perceived benefits over the risks of

academic misconduct (Adilov et al., 2024; Nikolic et al., 2023). Given that students may use chatbots for cognitive offloading, education institutions could set clear use policies about how the students could do so ethically. For instance, education institutions could allow the use of chatbots for legitimate cognitive offloading tasks such as editing support (Nikolic et al., 2023), grammar checking, and idea generation (Crawford et al., 2023).

The study's findings reveal that educators frequently use software to detect AI-generated text to prevent chatbot misuse for assessments, despite such software's limited reliability and accuracy (Adilov et al., 2024; Alexander et al., 2023; Ibrahim, 2023). In addition to using software, educators also rely on human judgement for detection (Chaudhry et al., 2023; Lancaster, 2023; Mao et al., 2024). This practice is consistent with suggestions to train educators to more accurately identify AI-generated text (Alexander et al., 2023; Perkins et al., 2024). It appears that chatbots and detection software are competing to outdo one another. As chatbots improve their capabilities, this may become a recurring scenario in which they play catch-up with one another (Nikolic et al., 2023). Students who misuse chatbots, on the other hand, will adopt strategies like prompt engineering to evade detection (Gorichanaz, 2023; Nikolic et al., 2023; Perkins et al., 2024).

The study's findings highlight the challenges that chatbots may pose to student learning performance evaluation if students misuse chatbots to complete their assessments. It appears that students who misuse chatbots face the possibility of academic integrity violations (Lancaster, 2023; Nikolic et al., 2023; Perkins et al., 2024), and the misuse will also compromise the reliability and validity of assessments in measuring students' true learning performance (Chaudhry et al., 2023; Gorichanaz, 2023; Ibrahim, 2023). This casts further doubts on these students' learning performance in terms of achieving the learning objectives, meeting the learning outcomes, and whether they have indeed grasped essential higher-order thinking skills in their studies (Blanke, 2024; Chaudhry et al., 2023; Lancaster, 2023). As a result, the value of education and education institutions' reputation may be questioned (Alexander et al., 2023; Chaudhry et al., 2023; Crawford et al., 2023). Students who misuse chatbots may also gain an unfair advantage over their peers, if the misuse is undetected, which creates an unfair situation for student learning performance evaluation (Chaudhry et al., 2023; Perkins et al., 2024; Yeadon et al., 2023). Additionally, doubts or accusations that arise about the authenticity of the assessments would in turn affect the trust between students and lecturers, which is detrimental to student engagement (Mao et al., 2024).

The study's findings identify several mitigation strategies for overcoming the challenges arising from students misusing chatbots for assessments. Education institutions should establish clear academic integrity policies, including plagiarism and cheating, and use policies that clearly define acceptable and unacceptable use of chatbots (Gorichanaz, 2023; Lancaster, 2023; Perkins et al., 2024). Furthermore, it is critical to effectively communicate these policies to educators and students (Perkins, 2023) through training and education (Alexander et al., 2023; Gorichanaz, 2023; Mao et al., 2024). It is also vital to engage key stakeholders in the education system to better understand their perspectives while formulating policies (Alexander et al., 2023; Bai et al., 2024; Chaudhry et al., 2023). This study's findings also indicate that, given the capabilities of chatbots, there is a need to redesign assessments into skill-based or authentic formats to improve their reliability and validity for measuring student learning (Blanke, 2024; Gorichanaz, 2023; Yeadon et al., 2023). On top of that, there is also a need to redesign learning to increase emphasis on higher-order thinking skills to better equip students with essential AI competences for the future (Chaudhry et al., 2023; Kortemeyer, 2023; Nikolic et al., 2023). This finding supports Guo and Lee's (2023) and Tlili et al.'s (2023) assertions that, with proper learning design, chatbots can assist students in developing thinking skills.

In conclusion, the thematic analysis synthesises and categorises the discussions in the articles into five major themes, each with respective sub-themes, and identifies the relationships between these themes. Students find reasons and benefits for adopting chatbots to assist them with their assessments, and they may employ counter-detection measures to prevent detection in the case of chatbot misuse. However, the integrity of the education system is challenged when chatbot misuse undermines the reliability and validity of assessments for performance measurement, resulting in the perception that the assessments may not accurately reflect students' true learning performance (Kooli, 2023). The education system must address these challenges with effective mitigation strategies rather than depend on detection strategies that fail to address the underlying challenges (Khalil and Er, 2023). The study's findings will help educational stakeholders better understand key discussions on the use of chatbots in student assessment, specifically the challenges that chatbot misuse may pose to student learning performance evaluation, and the various mitigation strategies to address these challenges.

Future research may consider two directions. First, broaden the scope of the thematic analysis by including articles from a wider range of databases to corroborate the themes identified in this study or uncover new

themes. Secondly, use the themes identified in this study as a framework to apply other research methods, such as questionnaire surveys or experiments, to triangulate the findings.

## References

- Adilov, N., Cline, J.W., Hanke, H., Kauffman, K., Meneau, L., Resendez, E., . . . Suntornpithug, N. (2024) "ChatGPT and the Course Vulnerability Index", *Journal of Education for Business*, Vol. 99 No. 2, pp 125-132.
- Al-Thani, S.J., Abdelmoneim, A., Daoud, K., Cherif, A. and Moukarzel, D. (2014) "A Perspective on Student Learning Outcome Assessment at Qatar University", *Quality in Higher Education*, Vol. 20 No. 3, pp 255-271.
- Alexander, K., Savvidou, C. and Alexander, C. (2023) "Who Wrote this Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education", *Teaching English with Technology*, Vol. 23 No. 2, pp 25-43.
- Alshurafat, H., Al Shbail, M.O., Hamdan, A., Al-Dmour, A. and Ensour, W. (2024) "Factors Affecting Accounting Students' Misuse of ChatGPT: An Application of the Fraud Triangle Theory", *Journal of Financial Reporting & Accounting*, Vol. 22 No. 2, pp 274-288.
- Altun, O., Saydam, M.B., Karatepe, T. and Dima, Ş.M. (2024) "Unveiling ChatGPT in Tourism Education: Exploring Perceptions, Advantages and Recommendations from Educators", *Worldwide Hospitality and Tourism Themes*, Vol. 16 No. 1, pp 105-118.
- Ansari, A.N., Ahmad, S. and Bhutta, S.M. (2023) "Mapping the Global Evidence around the Use of ChatGPT in Higher Education: A Systematic Scoping Review", *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-12223-4>
- Ayanwale, M.A. and Ndlovu, M. (2024) "Investigating Factors of Students' Behavioral Intentions to Adopt Chatbot Technologies in Higher Education: Perspective from Expanded Diffusion Theory of Innovation", *Computers in Human Behavior Reports*, Vol. 14, p 100396.
- Bai, J.Y.H., Zawacki-Richter, O. and Muskens, W. (2024) "Re-examining the Future Prospects of Artificial Intelligence in Education in Light of the GDPR and ChatGPT", *Turkish Online Journal of Distance Education*, Vol. 25 No. 1, pp. 20-32.
- Biggs, J.B. (2014) "Constructive Alignment in University Teaching", *HERDSA Review of Higher Education*, Vol. 1, pp 5-22.
- Blanke, J.M. (2024) "ChatGPT: The Sky is Not Falling", *Journal of Legal Studies Education*, Vol. 41 No. 1, pp 39-47.
- Boyatzis, R.E. (1998) *Transforming Qualitative Information: Thematic Analysis and Code Development*, SAGE, Thousand Oaks.
- Braun, V. and Clarke, V. (2006) "Using Thematic Analysis in Psychology", *Qualitative Research in Psychology*, Vol. 3 No. 2, pp 77-101.
- Braun, V. and Clarke, V. (2021) "One Size Fits All? What Counts as Quality Practice in (Reflexive) Thematic Analysis?", *Qualitative Research in Psychology*, Vol. 18 No. 3, pp 328-352.
- Chaudhry, I.S., Sarwary, S.A.M., El Refae, G.A. and Chabchoub, H. (2023) "Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT -- A Case Study", *Cogent Education*, Vol. 10 No. 1. <https://doi.org/10.1080/2331186X.2023.2210461>
- Crawford, J., Cowling, M. and Allen, K. (2023) "Leadership is Needed for Ethical ChatGPT: Character, Assessment, and Learning Using Artificial Intelligence (AI)", *Journal of University Teaching & Learning Practice*, Vol. 20 No. 3. <https://doi.org/10.53761/1.20.3.02>
- Crompton, H. and Burke, D. (2024) "The Educational Affordances and Challenges of ChatGPT: State of the Field", *TechTrends*, Vol. 68 No. 2, pp 380-392.
- Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., ... Wright, R. (2023) "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*, Vol. 71. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Elkhatat, A.M. (2023) "Evaluating the Authenticity of ChatGPT Responses: A Study on Text-Matching Capabilities", *International Journal for Educational Integrity*, Vol. 19. <https://doi.org/10.1007/s40979-023-00137-0>
- Farrokhnia, M., Banihashem, S.K., Noroozi, O. and Wals, A. (2024) "A SWOT Analysis of ChatGPT: Implications for Educational Practice and Research", *Innovations in Education and Teaching International*, Vol. 61 No. 3, pp 460-474.
- Flores, M.A., Veiga Simão, A.M., Barros, A. and Pereira, D. (2015) "Perceptions of Effectiveness, Fairness and Feedback of Assessment Methods: A Study in Higher Education", *Studies in Higher Education*, Vol. 40 No. 9, pp 1523-1534.
- Gao, X., Li, P., Shen, J. and Sun, H. (2020) "Reviewing Assessment of Student Learning in Interdisciplinary STEM Education", *International Journal of STEM Education*, Vol. 7 No. 1, pp 1-14.
- Gerritsen-van Leeuwenkamp, K.J., Joosten-ten Brinke, D. and Kester, L. (2019) "Students' Perceptions of Assessment Quality Related to their Learning Approaches and Learning Outcomes", *Studies in Educational Evaluation*, Vol. 63, pp 72-82.
- Gorichanaz, T. (2023) "Accused: How Students Respond to Allegations of Using ChatGPT on Assessments", *Learning: Research and Practice*, Vol. 9 No. 2, pp 183-196.
- Guo, Y. and Lee, D. (2023) "Leveraging ChatGPT for Enhancing Critical Thinking Skills", *Journal of Chemical Education*, Vol. 100 No. 12, pp 4876-4883.
- Ibrahim, K. (2023) "Using AI-based Detectors to Control AI-assisted Plagiarism in ESL Writing: "the Terminator Versus the Machines"", *Language Testing in Asia*, Vol. 13. <https://doi.org/10.1186/s40468-023-00260-2>

- Khalil, M. and Er, E. (2023) "Will ChatGPT Get You Caught? Rethinking of Plagiarism Detection", In: Zaphiris, P., Ioannou, A. (eds) *Learning and Collaboration Technologies*, Springer, Cham. [https://doi.org/10.1007/978-3-031-34411-4\\_32](https://doi.org/10.1007/978-3-031-34411-4_32)
- Kooli, C. (2023) "Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions", *Sustainability*, Vol. 15 No. 7, p 5614.
- Kortemeyer, G. (2023) "Could an Artificial-Intelligence Agent Pass an Introductory Physics Course?", *Physical Review Physics Education Research*, Vol. 19 No. 1. <https://doi.org/10.1103/PhysRevPhysEducRes.19.010132>
- Lancaster, T. (2023) "Artificial Intelligence, Text Generation Tools and ChatGPT -- Does Digital Watermarking Offer a Solution?", *International Journal for Educational Integrity*, Vol 19. <https://doi.org/10.1007/s40979-023-00131-6>
- Lau, A.M.S. (2016) "'Formative Good, Summative Bad?' - A Review of the Dichotomy in Assessment Literature", *Journal of Further and Higher Education*, Vol. 40 No. 4, pp 509-525.
- Mao, J., Chen, B. and Liu, J.C. (2024) "Generative Artificial Intelligence in Education and its Implications for Assessment", *TechTrends*, Vol. 68 No. 1, pp 58-66.
- Neto, J., Neto, F. and Furnham, A. (2023) "Predictors of Students' Preferences for Assessment Methods", *Assessment and Evaluation in Higher Education*, Vol. 48 No. 4, pp 556-565.
- Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G.M., Grundy, S., . . . Sandison, C. (2023) "ChatGPT Versus Engineering Education Assessment: A Multidisciplinary and Multi-Institutional Benchmarking and Analysis of this Generative Artificial Intelligence Tool to Investigate Assessment Integrity", *European Journal of Engineering Education*, Vol. 48 No. 4, pp 559-614.
- Niloy, A.C., Akter, S., Sultana, N., Sultana, J. and Rahman, S.I.U. (2024) "Is ChatGPT a Menace for Creative Writing Ability? An Experiment", *Journal of Computer Assisted Learning*, Vol. 40 No. 2, pp 919-930.
- Nugroho, A., Andriyanti, E., Widodo, P. and Mutiaraningrum, I. (2024) "Students' Appraisals Post-ChatGPT Use: Students' Narrative after Using ChatGPT for Writing", *Innovations in Education and Teaching International*. <https://doi.org/10.1080/14703297.2024.2319184>
- O'Neill, G. and Padden, L. (2022) "Diversifying Assessment Methods: Barriers, Benefits and Enablers", *Innovations in Education and Teaching International*, Vol. 59 No. 4, pp 398-409.
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., . . . Moher, D. (2021) "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews", *BMJ*, Vol. 372. <https://doi.org/10.1136/bmj.n71>
- Perkins, M. (2023) "Academic Integrity Considerations of AI Large Language Models in the Post-Pandemic Era: ChatGPT and Beyond", *Journal of University Teaching and Learning Practice*, Vol. 20 No. 2. <https://doi.org/10.53761/1.20.02.07>
- Perkins, M., Roe, J., Postma, D., McGaughran, J. and Hickerson, D. (2024) "Detection of GPT-4 Generated Text in Higher Education: Combining Academic Judgement and Software to Identify Generative AI Tool Misuse", *Journal of Academic Ethics*, Vol. 22 No. 1, pp 89-113.
- Tang, C.M. and Chaw, L.Y. (2023) "What Have People Discussed about ChatGPT in Malaysian Education? A Qualitative Content Analysis of News Articles", In: S.J. Johnston & S. Singh, eds. *22nd European Conference on e-Learning (ECEL 2023)*. Pretoria, South Africa, 26-27 October 2023. Reading: Academic Conferences International Limited.
- Tlili, A., Shehata, B., Adarkwah, M.A., Bozkurt, A., Hickey, D.T., Huang, R. and Agyemang, B. (2023) "What if the Devil is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education", *Smart Learning Environments*, Vol. 10 No 1, pp 1-24.
- Yeadon, W., Inyang, O., Mizouri, A., Peach, A. and Testrow, C. P. (2023) "The Death of the Short-Form Physics Essay in the Coming AI Revolution", *Physics Education*, Vol. 58 No. 3. <https://doi.org/10.1088/1361-6552/acc5cf>