2024

EJEL Volume 22, Issue 6

Editors

Paul Griffiths and Marija Cubric

Published by Academic Publishing International Limited
Curtis Farm, Kidmore End, Nr Reading, RG4 9AY, United Kingdom
karen.harris@academic-publishing.org

eISSN: 1479-4403

EJEL Volume 22, Issue 6

Special Issue: Artificial Intelligence (AI) in Education Part 2

Contents

Rubric Development and Validation for Assessing Tasks' Solving via AI Chatbots

Mohammad Hmoud[1], Hadeel Swaity[1], Eman Anjass[2] and Eva María Aguaded-Ramírez[2]

[1]Faculty of Educational Sciences, An-Najah National University, Nablus, Palestine

[2]Faculty of Educational Sciences, University of Granada, Spain

hmoud.tech@gmail.com (Corresponding author)

hadeelswaity17@gmail.com

emananjass@correo.ugr.es

eaguaded@ugr.es

https://doi.org/10.34190/ejel.22.6.3292

Abstract: This research aimed to develop and validate a rubric to assess Artificial Intelligence (AI) chatbots' effectiveness in accomplishing tasks, particularly within educational contexts. Given the rapidly growing integration of AI in various sectors, including education, a systematic and robust tool for evaluating AI chatbot performance is essential. This investigation involved a rigorous process including expert involvement to ensure content validity, as well as the application of statistical tests for assessing internal consistency and reliability. Factor analysis also revealed two significant domains, "Quality of Content" and "Quality of Expression", which further enhanced the construct validity of the evaluation scale. The results from this investigation robustly affirm the reliability and validity of the developed rubric, thus marking a significant advancement in the sphere of AI chatbot performance evaluation within educational contexts. Nonetheless, the study simultaneously emphasizes the requirement for additional validation research, specifically those entailing a variety of tasks and diverse AI chatbots, to further corroborate these findings. The ramifications of this research are profound, offering both researchers and practitioners engaged in chatbot development and evaluation a comprehensive and validated framework for the assessment of chatbot performance.

Keywords: Rubric development, AI chatbot, Validity, Tasks assessment, Educational technology

1. Introduction

The advent of artificial intelligence (AI) in the digital world has led to sweeping transformations, significantly impacting sectors such as automation, data processing, science research, and predictive analytics (García-Orosa, Canavilhas and Vázquez-Herrero, 2023; Yang, 2022). Among the various facets of AI, chatbots have attracted substantial academic interest due to their intricate algorithms and multifunctional abilities (Kooli, 2023). However, a significant impediment remains: the evaluation of these chatbots is complicated by the lack of a universally accepted and rigorously defined evaluation rubric. The rapid progress and integration of AI chatbots into educational settings have surpassed the development of systematic assessment approaches. This has created a significant gap in empirically evaluating their effectiveness in educational contexts. Addressing this gap requires recognizing the potential of AI chatbots in education and developing and validating assessment rubrics tailored to evaluate their performance accurately.

Recent theoretical developments have revealed that rubrics have become increasingly popular in educational evaluation, serving as assessment tools in various fields, such as appraising the quality of research publications in the medical sector (Moore, Bonnett, and Colbert-Getz, 2021). They also appear essential in verifying the authenticity of content in educational syllabi (Gregori-Giralt and Menéndez-Varela, 2019). These instances highlight the versatility and efficacy of rubrics in diverse educational scenarios. However, the rapid progress and integration of AI chatbots into educational settings have surpassed the development of systematic assessment approaches. This has created a significant gap in empirically evaluating their effectiveness in educational contexts (Smutný and Schreiberova, 2020). Task-solving assessment rubrics provide clear guidelines and quality indicators, which are pivotal to higher education for evaluating student performance across various tasks and assignments (McMurtrie, 2023; Tate, 2023). These allow assessors to measure the quality of students' work and provide valuable feedback (Bradley, Anderson, and Eagle, 2020).

Recent work in the field of AI demonstrated that the increasing prevalence of AI chatbots in education has sparked interest in their potential use for task-solving assessment (McMurtrie, 2023). For instance, AI chatbots like ChatGPT, developed by OpenAI, have demonstrated their potential in assisting students in task-solving

Cite this article: Hmoud, M. et al. 2024. "Rubric Development and Validation for Assessing Tasks' Solving via AI Chatbots", *Electronic Journal of e-Learning*, 22(6), pp 01-17, https://doi.org/10.34190/ejel.22.6.3292

processes such as drafting outlines, revising content, proofreading, and post-writing reflection (Su, Lin, and Lai, 2023; Vicente-Yagüe-Jara et al., 2023).

Based on the short preview above, the primary goal of this study is to fill the current research lacuna by developing a specialized task-solving assessment rubric to evaluate AI chatbots within an educational context. Discussions regarding the extent and boundaries of AI, amplified by recent advancements in machine learning and neural networks, remain contentious (Linardatos et al., 2020). Ongoing debates concern the differentiating factors between human and artificial intelligence and the restrictions of artificiality (Korteling et al., 2021). Despite AI's significant educational potential, it remains predominantly underexplored and undervalued, leading to its metaphorical description as the "Cinderella of the AI story" (Lameras and Arnab, 2021). Concerns regarding data privacy and the skepticism surrounding technology as a panacea have hampered the full integration of AI into mainstream education (Akgun and Greenhow, 2021; Flores-Vivar and García-Peñalvo, 2023).

Nevertheless, the assessment of AI, particularly chatbots, has emerged as a key area of inquiry (Maroengsit et al., 2019). However, The present evaluation methods lack a cohesive framework encompassing all the requisite elements for an exhaustive review (Gregori-Giralt and Menéndez-Varela, 2019). Prior studies on AI chatbots have emphasized the necessity of attributes such as relevance, accuracy, coherence, thoroughness, and efficiency (Moore, Bonnett, and Colbert-Getz, 2021). However, a rubric that integrates these elements comprehensively and uniformly is conspicuously absent from the current literature. By examining the integration of AI chatbots into task-solving assessment, this study aims to identify best practices for educators to effectively incorporate this technology while preserving the authenticity of student work. Thus, this research is driven by the potential benefits of integrating AI chatbots into task-solving assessment in higher education. Consequently, through this exploration of the effectiveness of AI chatbots and the creation of suitable rubrics, it is evident that the overarching goal of the study, in addition to the specific objectives laid out, is to enhance the task-solving assessment process and provide valuable insights for educators and researchers in the field.

1.1 Utilizing Chatbots in Higher Education for Enhanced Learning and Task Solving

The Chatbots have been influential in the field of education because of their significant impact. The digital transformation ushered in by artificial intelligence (AI) has profoundly impacted numerous sectors, with notable effects in automation, data processing, scientific research, and predictive analytics. By the same token, chatbots have garnered significant academic interest in this technological revolution for their complex algorithms and versatile functionalities, especially in the educational sector (García-Orosa, Canavilhas and Vázquez-Herrero, 2023; Kooli, 2023; Yang, 2022). Recent studies by Kim and Lee (2023) and Hmoud et al. (2024) have shown growing appeal, which further highlights the transformative potential of integrating AI chatbots into educational practices. Kim and Lee (2023) delve into Student AI Collaboration (SAC) and its influence on creative tasks, finding that SAC notably enhances creativity, expressivity, and task effectiveness, with the degree of impact influenced by students' attitudes towards AI and their drawing skills. This underlines the importance of adaptive scaffolding in educational AI systems to accommodate diverse student needs, thus improving the learning experience through personalized support.

Building on these insights, Hmoud et al. (2024) examine the motivational dimensions of ChatGPT usage in learning environments. They identify significant effects on student motivation across five core areas: task enjoyment, reported effort, result assessment, perceived relevance, and interaction. Their research notably emphasizes ChatGPT's ability to amplify task enjoyment, indicating that interacting with AI chatbots can greatly enhance students' satisfaction and curiosity, which in turn improves task motivation. However, they also caution about challenges concerning the accuracy of information provided by chatbots, highlighting the essential role of critical evaluation skills among students.

One of the major topics to be investigated in this field is the effect of AI on pedagogical tools used by teachers and educators. Further extending the discourse on the utility of AI in education, Baidoo-Anu and Ansah et al. (2023) illuminate the role of generative AI chatbots as effective pedagogical tools. These chatbots offer conversational assistance, support multiple communication modes, and provide multilingual capabilities, making them cost-effective, scalable, and seamlessly integrated with existing educational technologies. Their ability to offer data analytics and insights enables educators to refine teaching methodologies, showcasing the multifaceted benefits of AI chatbots in enhancing pedagogical practices (Hmoud and Shaqour, 2024). Supporting this viewpoint, Ilieva et al. (2023) highlight the invaluable role of AI chatbots in higher education, especially in providing personalized assistance in advanced and specialized subjects. By promoting selfdirected and independent learning, and facilitating access to scholarly resources, these chatbots significantly support students in their research activities, including assistance with literature reviews and research methodology guidance, thereby fostering academic research and enhancing learning autonomy.

Amidst these technological advancements, a series of recent studies concluded that debates surrounding the distinctions between human and artificial intelligence and concerns about technology as a universal remedy continue to be contentious. However, minor issues have been experienced. Issues such as data privacy and skepticism towards the wholesale integration of AI into mainstream education further complicate the landscape (Akgun and Greenhow, 2021; Flores-Vivar and García-Peñalvo, 2023). Despite these challenges, assessing AI, especially chatbots, in educational contexts has become an essential area of inquiry for many scholars. Yet, the lack of a cohesive framework for comprehensive evaluation points to a significant gap in the literature (Gregori-Giralt and Menéndez-Varela, 2019; Maroengsit et al., 2019).

Based on the preceding, the present work aims to address this gap by advocating for the development of systematic approaches to evaluate the effectiveness of AI chatbots in educational settings, particularly in enhancing task-solving and learning experiences. It highlights the urgent need for research focused on creating reliable methodologies for assessing AI chatbots, thereby contributing to the refinement of their integration into educational frameworks (Smutný and Schreiberova, 2020).

The contributions of Kim and Lee (2023), Hmoud et al. (2024), Baidoo-Anu and Ansah et al. (2023), and Ilieva et al. (2023) offer a nuanced perspective on the implications of using chatbots for task-solving in higher education. They suggest that AI chatbots can significantly enrich student learning experiences by fostering creativity, motivation, and engagement, albeit with an acknowledgment of their limitations. These findings advocate for the development of educational AI that is both adaptable and responsive, capable of supporting diverse learning activities while encouraging critical engagement with content. This comprehensive approach positions chatbots like ChatGPT as invaluable tools in advancing higher education, provided their application is balanced with thoughtful instructional design and rigorous evaluation practices.

1.2 Developing a Rubric for Task-Solving Assessment Through Chatbots: Implications for Teaching, Learning, and Assessment Practices

Integrating rubrics in evaluating tasks facilitated by chatbot technology, such as AI chatbots, is becoming increasingly vital in educational contexts. Rubrics, as structured assessment tools, offer clear and explicit criteria that significantly enhance the evaluation of student work (Brookhart, 2018; Tan, 2020). They also can provide specific standards and expectations for assessing student performance during interactions with chatbots (Bradley, Anderson, and Eagle, 2020). This approach improves students' understanding of assessment criteria and promotes a more objective and consistent assessment process (El-Magd, 2022; Tan, 2020). Furthermore, rubrics facilitate formative feedback and enhance metacognitive skills, guiding students to understand better and meet assignment expectations (De Vera, 2023; Panadero and Jonsson, 2020).

A series of recent studies have elucidated that while rubrics are widely recognized for their benefits in educational evaluation, the absence of a universally accepted evaluation rubric for AI chatbots highlights a significant research gap. (Gregori-Giralt and Menéndez-Varela, 2019; Ilieva et al., 2023; Kooli, 2023; Moore, Bonnett, and Colbert-Getz, 2021; Smutný and Schreiberova, 2020). To address this gap, recent studies by Almasre (2024), Cope, Kalantzis, and Searsmith (2021), and Abbas, Jam, and Khan (2024) have shed light on both the potential enhancements and challenges of incorporating AI into educational assessments. For instance, Almasre (2024) and Cope, Kalantzis, and Searsmith (2021) emphasize AI's capacity to revolutionize educational assessments and facilitate diverse learning pathways through formative assessments and innovative feedback mechanisms. Conversely, Abbas, Jam, and Khan (2024) caution against the potential adverse effects of excessive AI chatbot usage, such as procrastination and diminished academic performance, highlighting the necessity for a balanced integration of AI.

This literature emphasizes the need to develop a Task-Solving Assessment Rubric tailored specifically for AI Chatbots to bolster higher education teaching, learning, and assessment practices. This kind of rubric would amalgamate dynamic evaluation criteria, encompassing relevance, accuracy, and efficiency, thereby establishing a standardized framework for appraising the contributions of AI chatbots within educational settings (Bradley, Anderson, and Eagle, 2020; Lim, 2022; McMurtrie, 2023; Su, Lin, and Lai, 2023; Tate, 2023; Vicente-Yagüe-Jara et al., 2023).

This initiative aims not only to address a noticeable gap in the current literature but also to furnish a structured evaluation tool that resonates with the innovative capabilities of AI chatbots, thereby fostering a conducive and efficient learning environment. As higher education institutions adapt to the evolving landscape of AI applications, the development and validation of such a rubric will play a pivotal role. It will ensure the integration and utilization of AI chatbots in a manner that nurtures academic advancement and aligns with desired learning outcomes. This is highlighted by studies from El-Magd (2022), Panadero and Jonsson (2020), De Vera (2023), and Tenakwah et al. (2023).

1.3 Conceptual Framework

The conceptual framework for developing a Task-Solving Assessment Rubric for AI Chatbots in higher education seeks to fill a notable gap in existing research, aiming to enhance educational practices by systematically evaluating AI chatbots. This framework is intricately designed around the principles of Competency-Based Learning (CBL) (Henri, Johnson, and Nepal, 2017; Tenakwah et al., 2023) and informed by Brown's (2012) methodology for rubric development, reflecting a comprehensive approach to assessing AI chatbot interactions within educational settings. Central to this framework is establishing explicit goals to define essential competencies required for effective engagement with AI chatbots. These competencies encompass knowledge, skills, and abilities critical to navigating AI technologies, ensuring students are equipped for task-solving activities facilitated by chatbots. The design phase of the rubric, guided by the CBL framework, presents educators with a choice between analytic and holistic assessment methods, emphasizing mastery over critical competencies. This phase is pivotal in structuring rubric categories and scoring ranges that quantitatively measure competency attainment, thereby offering clear and actionable feedback to enhance learning outcomes (Brown, 2012; Henri, Johnson, and Nepal, 2017).

Implementing the rubric involves introducing it to students as a preparatory tool, embodying the concept of preemptive feedback. This approach aligns student efforts with the competencies outlined, setting a clear expectation before task engagement. The subsequent evaluation of student work employs this rubric to provide targeted feedback, highlighting strengths and identifying improvement areas, facilitating a nuanced development of competencies (Brown, 2012). Evaluating the rubric's effectiveness extends beyond its application, encompassing an analysis of its reliability, fairness, validity, and usability. This evaluation is instrumental in refining the rubric, ensuring its adaptability and relevance in the face of rapidly evolving AI technologies. The integration of Brown's structured development process with the CBL approach underpins the framework's robustness, enhancing the precision and utility of the rubric as an educational tool. It emphasizes the role of explicit learning objectives and feedback mechanisms, fostering an environment conducive to effectively applying knowledge and skills in real-world scenarios (Brown, 2012; Henri, Johnson, and Nepal, 2017).

Based on studies by Almasre (2024), Cope, Kalantzis, and Searsmith (2021), Abbas, Jam, and Khan (2024), and others, the framework suggests a balanced use of AI in education. It suggests that AI's shoud be used for innovations while avoiding any negative effects on student engagement and performance. This balanced integration aims to support academic growth and foster a positive learning environment, thereby contributing to the broader educational discourse on AI's role in enhancing learning outcomes and competency development (Bradley, Anderson, and Eagle, 2020; De Vera, 2023; El-Magd, 2022; Gregori-Giralt and Menéndez-Varela, 2019; Ilieva et al., 2023; Kooli, 2023; Lim, 2022; McMurtrie, 2023; Moore, Bonnett, and Colbert-Getz, 2021; Panadero and Jonsson, 2020; Smutný and Schreiberova, 2020; Su, Lin, and Lai, 2023; Tate, 2023; Vicente-Yagüe-Jara et al., 2023).

This study's objectives—developing, validating, and applying a Task-Solving Assessment Rubric for AI Chatbots—underscore the imperative to evaluate AI's educational effectiveness systematically. This research contributes significantly to understanding how AI chatbots can augment task-solving assessment processes in higher education by formulating rubrics that encapsulate essential evaluation elements and align with competency-based learning outcomes. This endeavor seeks to address a critical research void and ensure that AI chatbot integration into educational settings is effective and conducive to fostering critical thinking and originality among students (Lim, 2022).

2. Methodology

An instrumental and descriptive study of validity and reliability of a rubric was carried out. The descriptive assessment measures, often known as rubrics, are among the major current instruments associated with this tendency. These rubrics are based on a graded set of rules that are employed in complete holistic evaluation, with a large capacity for standardizing the assessment of students' performance levels. As a result, they improve the validity and reliability of performance evaluation, resulting in improved judgment, evaluation, and identification of students' strengths and weaknesses (Stiggins, 1997). Rubrics are indicators that offer explicit and obvious standards for evaluating specific talents or tasks (Reddy and Andrade, 2010; Stanley, 2021b). They are made up of a set of performance indicators and predefined levels of achievement that allow for a systematic and consistent evaluation procedure (Brookhart, 2013). Furthermore, rubrics allow teachers to break down complex skills into individual components, giving students comprehensive feedback and highlighting their strengths and areas for improvement (Panadero and Jonsson, 2013; Stanley, 2021b). The procedure for creating and developing an assessment rubric for solving tasks with chatbots powered by artificial intelligence are explained below:

Phase 1: Identifying the objective of constructing the rubric:

The current task's purpose is to use the assessment rubric as a tool to review and create the criteria that should be met by tasks solved using AI-supported chatbots. As a result, it tries to improve the level and efficiency of these jobs while identifying the strengths and shortcomings in the responses and comments supplied by these platforms (Brookhart, 2018; El-Majd, 2022; Tan, 2020).

Phase 2: Selecting the type of rubric and justifying your choice:

We chose an Analytic Descriptive Rubric after researching educational literature on the design and development of assessment tools. This rubric type was chosen because of its precision, objectivity, realism, and comprehensiveness. It is an alternate assessment technique that focuses on the performance of learners, covering processes and outcomes. It is based on qualitative performance assessment using descriptive rating scales and gives information about the strengths and weaknesses of the many dimensions and components of performance. This data can be used to improve performance by giving precise and detailed feedback to both teachers and students, hence aiding the teaching, and learning processes (Hack, 2015).

Phase 3: Initial Scale Description and Item Development:

The assessment scale's content was generated and formulated using educational literature, studies, and previous research on artificial intelligence and chatbots (Abdul-Kader and Woods, 2015; Brandtzaeg and Følstad, 2017; Hill, Ford, and Farreras, 2015; Jain et al., 2018; Luger and Sellen, 2016). Following that, a brainstorming session was held to discover relevant indications that support the scale.

In terms of item formation, they were expressed in a clear procedural form to promote observation and comprehension by both teachers and students. The pieces were written in short, succinct phrases that have only one meaning. Creating the Assessment Scale's Structure and Organization: The scale's basic form includes 37 criteria, which are as follows:

1. **Accuracy**: Did the chatbot provide a correct and accurate response to the user's query or request?
2. **Relevance**: How relevant was the chatbot's response to the user's question or comment?
3. **Efficiency**: Did the chatbot's response fully answer the user's question or was additional clarification required?
4. **Clarity**: Was the chatbot's response clear and easy to understand, or was it confusing or ambiguous?
5. **Context-Awareness**: Did the chatbot appropriately consider the context of the conversation when providing a response?
6. **Conversational Flow**: Did the chatbot's response fit naturally within the flow of the conversation, or did it disrupt the dialogue?
7. **Empathy**: Did the chatbot respond with an appropriate level of empathy and emotional understanding?
8. **Politeness**: Did the chatbot demonstrate politeness and respect towards the user in its responses?
9. **Speed of Response**: How quickly did the chatbot provide a response?
10. **Coherence**: Were the chatbot's responses consistent throughout the conversation, and did it maintain a coherent line of thought?
11. **Grammar and Spelling**: Were the chatbot's responses free of grammar mistakes and spelling errors?
12. **Personalization**: Did the chatbot personalize its responses based on the user's specific needs and preferences?
13. **Engagement**: Did the chatbot's responses engage the user and promote further conversation?
14. **Error Handling**: How well did the chatbot handle misunderstandings or errors in the user's inputs?
15. **Fallback Strategy**: Was the chatbot able to handle unrecognized inputs or queries gracefully?

16. **Escalation Process**: How effectively did the chatbot hand off the conversation to a human agent when it was unable to assist the user?
17. **Adherence to Guidelines**: Did the chatbot adhere to pre-set guidelines (such as not providing medical, legal, or financial advice unless specifically trained and authorized to do so)?
18. **Security and Privacy**: Did the chatbot properly handle user data, ensuring its security and privacy?
19. **Multilingual Capability**: Can the chatbot effectively communicate in multiple languages as per user requirements?
20. **Appropriateness of Language**: Does the chatbot perfectly appropriate for the task at hand.
21. **Information Verification**: Does the chatbot confirm the accuracy of information provided by the user when necessary?
22. **Domain Knowledge**: How well does the chatbot respond to queries that are specific to the domain it is designed for?
23. **Handling of Complex Queries**: Can the chatbot handle complex queries, or does it only manage simple, straightforward questions?
24. **Self-Correction**: Can the chatbot identify when it's made an error and correct it in real-time?
25. **User Feedback Mechanism**: Does the chatbot have a mechanism to receive and incorporate user feedback?
26. **Simplicity**: Is the chatbot easy to interact with, even for users who aren't very tech-savvy?
27. **Up to date**: Does the chatbot provide responses that are current and up-to-date, especially for time-sensitive or dynamic information?
28. **Scalability**: Can the chatbot handle a large volume of conversations simultaneously without a drop in performance?
29. **Usability**: Is the chatbot user-friendly? Does it offer an intuitive interface?
30. **Informativeness**: Does the chatbot provide a sufficient amount of detail in its responses, without overwhelming the user with unnecessary information?
31. **Adaptability**: Can the chatbot learn from past interactions and improve its responses over time?
32. **Response Diversity**: Does the chatbot vary its responses to avoid sounding too robotic or repetitive?
33. **Comprehensiveness**: Does the chatbot covering all aspects of the task in detail?
34. **Crisis Management**: How effectively does the chatbot manage crisis situations or urgent user needs?
35. **Argument and Evidence**: Does the essay present a clear and compelling argument, and is this argument supported by substantial, reliable evidence from credible sources?
36. **Language and Tone**: Does the essay use appropriate, sophisticated, and consistent language throughout, and does it maintain an academic tone suitable for the context of the assignment?
37. **Creativity and Originality**: Does the essay provide unique insights or original perspectives, and does it demonstrate innovative thinking or creativity in its approach to the topic?

Phase 4: Choosing Assessment of Performance Levels:

To determine the performance level of each facet within each axis, the scale consists of five progressive standard and descriptive levels. Performance scores, ranging from 1 to 5, are assigned based on the extent to which performance indicators are met by the student, with higher scores indicating a better level of standard achievement (Chi, 2013; Jonsson and Svingby, 2007; Stanley, 2021).

Phase 5: Instructions for Creating a Scale:

The scale starts with instructions for the evaluator, which include a description of the scale's structure and components, the levels of performance assessment, and an explanation of the evaluator's responsibilities. An illustrated example of utilizing the scale to evaluate performance outcomes is provided and will be shown later (Chi, 2013; Stanley, 2021).

Phase 6: Verification of Rubric Scale Validity:

The initial version of the assessment scale was provided to 12 professional reviewers working in the field of computing and information technology to determine the necessity for any revisions or alterations to the content of the assessment scale. A Microsoft Form questionnaire was emailed to them in order to solicit their feedback and suggestions on each criterion. For making decisions, the following scale was used: (1) Criterion is unneeded and inappropriate, (2) Criterion is valuable but unnecessary, (3) Criterion is necessary and appropriate, and (4) Criterion is necessary and appropriate. The questionnaire also includes an open-ended place for reviewers to add any relevant remarks (de La Rosa Gómez, Cano, and Diaz, 2019).

Following receipt of the reviewers' replies on the assessment scale criteria, the Content Validity Ratio (CVR) for each criterion was determined using the formula which appear in Equation 1.

$$\mathrm{CVR} = \frac{\mathrm{ne} - \frac{\mathrm{N}}{2}}{\frac{\mathrm{N}}{2}} \qquad (1)$$

Where *ne* is the number of reviewers who say the criterion is "necessary and appropriate", *N* represents the total number of reviewers.

The CVR value is zero when half of the reviewers indicate that the criterion is required and appropriate while the other half do not. The CVR value is 1 when all reviewers agree that the criterion is required and reasonable. The CVR value is between 0 and 0.99 when more than half of the reviewers indicate that the criterion is required and appropriate, but not all of them. CVR is a useful statistical measure for examining items and deciding whether to accept or reject them based on reviewers' decisions. It is widely accepted as a tool for determining content validity (Wilson, Pan, and Schumsky, 2012).

After gathering the reviewers' data, Microsoft Excel software was utilized to complete the essential statistical and mathematical calculations for analysis, such as CVR, CVI, and Kappa. Decisions on item acceptance or rejection were made based on the CVR ratio, which should be greater than 0.667 for each item, considering the number of reviewers (12), as mentioned in the study by Ayre and Scally (2014). By computing the Content Validity Index (I-CVI) for each item, criteria were omitted based on recommendations of a study by Mishra (2017). The following formula used to calculate the I-CVI as show in Equation 2.

$$\mathrm{ItemCVI} = \frac{\mathrm{ne}}{\mathrm{N}} \qquad (2)$$

Where *ne* number of reviewers designating the criterion as "necessary and appropriate", *N* is total number of reviewers.

To determine the consistency of the reviewers for each item, the modified kappa coefficient (*), commonly known as the inter-rater agreement strength, was calculated. Equation 3 was used to arrive to this conclusion:

$$\mathrm{k}^* = \frac{\mathrm{ItemCVI} - \mathrm{p_c}}{1 - \mathrm{p_c}} \qquad (3)$$

Where ItemCVI is the content validity index, and pc is the observed percentage of reviewer agreement.

To guarantee the accuracy and agreement of the reviewers, the criteria were subsequently eliminated based on the I-CVI values and the adjusted kappa coefficient (*). These statistical methods were used in accordance with accepted standards for assessment validity and reliability. Mishra's study (2017) offers valuable insights for refining assessment criteria, enhancing validity and ensuring a robust evaluation process.

Phase 7: Verification of Rubric Scale Reliability:

Two techniques were used to evaluate the rubric's reliability:

a) The reliability of the inter-rater agreement was assessed across all items using the Kappa coefficient calculation.

An-Najah University Graduate Studies in Education students were given an assignment to use the chatbots to analyze the advantages and disadvantages of a contemporary teaching approach based on the course's characteristics and requirements. Then, using an assessment tool created by five seasoned professors who acted as independent assessors, an evaluation procedure was carried out. Using the Interrater Reliability technique, this was done to confirm the reliability of the evaluation scale.

The following processes were done in order to determine the overall agreement percentage: a list of each assessor's evaluations for each item on the scale was prepared. Then, for each item, agreement points (1) and disagreement points (0) were determined. The Holsti equation (Holsti, 1970) was used to calculate the sum of agreement points and to obtain the percentage of agreement between the assessors' evaluations:

Agreement Percentage is calculated as follows = 5 * Number of Agreements / (Number of Items Assessed by Assessors 1 + Number of Items Assessed by Assessors 2 + … + Number of Items Assessed by Assessors 5) * 100%.

b) Prior to implementation, the reliability of the assessment scale was further validated using a method known as Intrarater reliability. Thirty students who participated in the previous evaluation were asked to retake the work when there was ample time in between the tests. The Test/Retest method is one way to assess the accuracy of measurement tools, despite some disadvantages including test familiarity and the potential impact of the students' earlier performance.

Phase 8: Apply Rubric:

The assessment scale was applied to 144 university students who studied at Graduate Studies in Education and completed a final task using chatbots during the second semester of the academic year 2022/2023 after reviewers and evaluators confirmed the validity of the assessment scale and established its reliability.

3. Results

Nine criteria were accepted after determining the Conversion Rate (CVR) for each criterion, as shown in the Table 1.

Table 1: Simplified Table of CVR

Criteria item	Num of Experts Agree an item	*CVR*	Result
Accuracy	12	1	Accept
Relevance	12	1	Accept
Efficiency	12	1	Accept
Clarity	9	0.5	Reject
Context-Awareness	9	0.5	Reject
Conversational Flow	8	0.33	Reject
Empathy	9	0.5	Reject
Politeness	8	0.33	Reject
Speed of Response	6	0	Reject
Coherence	12	1	Accept
Grammar and Spelling	12	1	Accept
Personalization	9	0.5	Reject
Engagement	9	0.5	Reject
Error Handling	8	0.33	Reject
Fallback Strategy	9	0.5	Reject
Escalation Process	8	0.33	Reject
Adherence to Guidelines	7	0.17	Reject
Security and Privacy0	6	0	Reject
Multilingual Capability	6	0	Reject
Appropriateness of Language	9	0.5	Reject
Information Verification	9	0.5	Reject
Domain Knowledge	9	0.5	Reject
Handling of Complex Queries	8	0.33	Reject
Self-Correction	8	0.33	Reject
User Feedback Mechanism	9	0.5	Reject
Simplicity	8	0.33	Reject
Up to date	8	0.33	Reject
Scalability	8	0.33	Reject
Usability	8	0.33	Reject
Informativeness	7	0.17	Reject
Adaptability	8	0.33	Reject

Criteria item	Num of Experts Agree an item	*CVR*	Result
Response Diversity	9	0.5	Reject
Comprehensiveness	11	0.83	Accept
Crisis Management	8	0.33	Reject
Argument and Evidence	10	0.67	Accept
Language and Tone	11	0.83	Accept
Creativity and Originality	12	1	Accept

All the accepted criteria clearly show a degree of agreement among the reviewers that may be regarded as almost flawless, as indicated by the previously stated indicators. When the I-CVI and the modified kappa coefficient for all the criteria are calculated. The values in accordance with the criterion number was shown in Table 2. Instead of calculating each item separately using CVR, the CVI indicator's aggregate result frequently results in a scale that is more effective overall. CVR is a practical statistical technique for assessing the validity of each individual item based on reviewers' evaluations. The total average CVR of all the elements that make up the instrument is represented numerically by the CVI, in contrast (Gilbert and Prion, 2016).

Table 2: Simplified Table of Items CVI, Kappa Coefficients

Criteria item	*I-CVI*	*pc*	*K**	Strength of Agreement
Accuracy	1	0	1	Almost Perfect
Relevance	1	0	1	Almost Perfect
Efficiency	1	0	1	Almost Perfect
Clarity	0.75	0.05	0.74	Substantial
Context-Awareness	0.75	0.05	0.74	Substantial
Conversational Flow	0.67	0.12	0.62	Substantial
Empathy	0.75	0.05	0.74	Substantial
Politeness	0.67	0.12	0.62	Substantial
Speed of Response	0.5	0.23	0.35	Fair
Coherence	1	0	1	Almost Perfect
Grammar and Spelling	1	0	1	Almost Perfect
Personalization	0.75	0.05	0.74	Substantial
Engagement	0.75	0.05	0.74	Substantial
Error Handling	0.67	0.12	0.62	Substantial
Fallback Strategy	0.75	0.05	0.74	Substantial
Escalation Process	0.67	0.12	0.62	Substantial
Adherence to Guidelines	0.58	0.19	0.48	Moderate
Security and Privacy0	0.5	0.23	0.35	Fair
Multilingual Capability	0.5	0.23	0.35	Fair
Appropriateness of Language	0.75	0.05	0.74	Substantial
Information Verification	0.75	0.05	0.74	Substantial
Domain Knowledge	0.75	0.05	0.74	Substantial
Handling of Complex Queries	0.67	0.12	0.62	Substantial
Self-Correction	0.67	0.12	0.62	Substantial
User Feedback Mechanism	0.75	0.05	0.74	Substantial
Simplicity	0.67	0.12	0.62	Substantial

Criteria item	*I-CVI*	*pc*	*K**	Strength of Agreement
Up to date	0.67	0.12	0.62	Substantial
Scalability	0.67	0.12	0.62	Substantial
Usability	0.67	0.12	0.62	Substantial
Informativeness	0.58	0.19	0.48	Moderate
Adaptability	0.67	0.12	0.62	Substantial
Response Diversity	0.75	0.05	0.74	Substantial
Comprehensiveness	0.92	0	0.92	Almost Perfect
Crisis Management	0.67	0.12	0.62	Substantial
Argument and Evidence	0.83	0.02	0.83	Almost Perfect
Language and Tone	0.92	0	0.92	Almost Perfect
Creativity and Originality	1	0	1	Almost Perfect

Therefore, after removing the items on which the reviewers differed, the CVI for the new evaluation scale tool was determined using Equation 4.

$$\text{CVI} = \frac{\sum \text{CVR}}{\text{retained numbers}} = \frac{8.33}{9} = 0.926 \qquad (4)$$

While Davis (1992) contends that a CVI value of 0.80 is ideal, the study by Tilden, Nelson, and May (1990) contends that CVI values should surpass 0.70. As a result, the final CVI value of 0.93 is higher than 0.8, demonstrating the validity, reliability, and applicability of the overall evaluation scale instrument. As a result, the scale was built with the marks, which are shown in Table 3.

Table 3: Rubric Assessment Scale

Criteria	(1) Point	(2) Points	(3) Points	(4) Points	(5) Points
Relevance	Fully and directly addresses all aspects of the task.	Directly addresses most aspects of the task, minor points missing.	Partially addresses the task but lacks important details.	Vaguely related but does not directly address the task.	Unrelated to the task.
Accuracy	Completely accurate.	Almost entirely accurate.	Mostly accurate but contains minor inaccuracies.	Contains a few factual inaccuracies.	Contains several factual inaccuracies.
Efficiency	Efficient and succinct.	Mostly efficient, with minor room for improvement.	Somewhat efficient, with room for improvement.	Could be more succinct or well-structured.	Unnecessarily lengthy or convoluted.
Coherence	Completely coherent and easy to follow.	Very coherent with only minor issues.	Mostly coherent, with a few confusing statements.	Some coherence but difficult to understand.	Largely incoherent or nonsensical.
Comprehensiveness	Extremely comprehensive, covering all aspects of the task in detail.	Quite comprehensive, only missing a few minor points.	Covers most of the task but lacks some details.	Covers a few elements of the task but misses many key points.	Barely touches on the task.
Grammar and Spelling	Excellent grammar and spelling with no errors.	Good grammar and spelling with a few minor errors.	Acceptable grammar and spelling, but with several mistakes.	Poor grammar and spelling with many errors.	Unacceptable grammar and spelling with frequent errors.

Criteria	(1) Point	(2) Points	(3) Points	(4) Points	(5) Points
Argument and Evidence	Clearly articulated, strong argument supported by substantial evidence.	Solid argument with adequate evidence.	Argument present but lacks substantial supporting evidence.	Weak argument with little or insufficient evidence.	Absent or unclear argument with no evidence.
Language and Tone	Highly sophisticated and nuanced language and tone.	Sophisticated language and tone with minor inconsistencies.	Acceptable language and tone but with some inconsistencies.	Inappropriate or inconsistent language and tone.	Poor language use and inappropriate tone.
Creativity and Originality	Unique, insightful, and innovative approach.	Somewhat unique with some insightful thoughts.	Ordinary approach with few insights.	Lack of originality, few insights.	Completely lacks creativity or originality.

According to Holsti (1970), an agreement percentage of 85% or higher indicates strong instrument reliability, whereas one of less than 70% suggests low instrument reliability. The agreement rate between the faculty assessors in grading the thirty students using the assessment scale was higher than 85%, and points to the assessment scale's Good Reliability.

Following the second evaluation of the students' performance, the results of both tests were exceeded 70%, showing high instrument reliability, the evaluation scale nonetheless maintained good reliability as a measurement tool (Streiner, 2003, p.102). The convergent validity of the scale's final form was investigated to ascertain the construct validity of the assessment scale. Exploratory factor analysis employing the principal component analysis approach and Varimax rotation was used to achieve this. Two domains with an eigenvalue greater than one was produced by the analysis, and it had 6 items (Accuracy, Relevance, Coherence, Comprehensiveness, Grammar and Spelling, Argument and Evidence) for the first domain which named "Quality of Content" and 3 items (Efficiency, Language and Tone, Creativity and Originality) for the second domain which named "Quality of Expression". Table 4 shows that the combined factors explained 84.3% of the overall variance.

Table 4: Total Variance Explained

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.437	71.525	71.525	4.498	49.978	49.978
2	1.150	12.777	84.302	3.089	34.324	84.302
3	.645	7.162	91.465			
4	.312	3.470	94.935			
5	.173	1.919	96.854			
6	.118	1.310	98.164			
7	.091	1.015	99.179			
8	.067	.741	99.920			
9	.007	.080	100.000			

Note: ***Extraction Method: Principal Component Analysis.***

According to Gorsuch's (2014) findings, a factor is considered valid when at least three different variables show factor loadings that are more than 0.3. The factor is regarded as insignificant or unimportant if it does not satisfy this requirement. The factor in question was confirmed and accepted in the current investigation because all of the assessment scale's items showed significant loadings on the factor. This supports the validity of the instrument, which is shown in Table 5.

Table 5: Rotated Component Matrixa

Criteria	Component	
	1	2
Accuracy	0.887	
Relevance	0.689	0.535
Efficiency	0.523	0.762
Coherence	0.835	0.408
Comprehensiveness	0.885	
Grammar and Spelling	0.869	
Language and Tone		0.908
Creativity and Originality		0.921
Argument and Evidence	0.778	0.457

Note: Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Kaiser-Meyer-Olkin (KMO) test, which assesses the suitability of the sample size for proving the efficacy of factor analysis and whether the partial correlations between variables are minimal, was carried out to ensure the effectiveness of conducting factor analysis on the instrument. Field (2018) claims that high values in the KMO test findings, which are greater than 0.7, indicate that factor analysis would be helpful for our data. After running the test, it was discovered that the KMO value was equal to 0.802, suggesting that the sample was adequate, and that factor analysis had been successful, as indicated in Table 6.

Table 6: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.802
Bartlett's Test of Sphericity	Approx. Chi-Square	259.401
	df	36
	Sig.	.000

In order to determine how much each item contributes to measuring the rubric scale as a whole, the Pearson correlation coefficient was determined for each item in respect to the overall rubric scale and that were more than 0.6, which indicates that the rubric as a whole scale is strong (Turney, 2022).

The internal consistency reliability of the estimate scale was examined using the SPSS software by determining the Cronbach's alpha coefficient, which was 0.948, and that indicates good internal Consistency for the Rubric Scale items according to Streiner (2003, p. 102). Also, another estimate scale's reliability was evaluated in SPSS using the split-half method for rubric scale items, which divided them into two equivalent halves using the ODD-EVEN technique, as shown in the table below. The results showed that the Robustness scale was reliable, with Cronbach's alpha coefficients of 0.905 and 0.883 for each half, respectively. Furthermore, the Spearman-Brown coefficient was 0.970, and the Guttman coefficient was 0.966, suggesting outstanding reliability for both sides of the Robustness scale.

4. Discussion

From the results above, key issues emerged that must be addressed. In the first place, the results demonstrated in this chapter match state of the art methods. The main promising finding is that integrating AI chatbots into educational settings marks an essential evolution in teaching and learning methodologies. With AI's capacity to revolutionize sectors from predictive analytics to scientific research (García-Orosa, Canavilhas and Vázquez-Herrero, 2023; Yang, 2022), its application in education promises to enhance teaching and student engagement. Recent studies show that AI chatbots can potentially improve task-solving assessments, as demonstrated by their ability to enhance creativity, motivation, and student engagement (Hmoud et al., 2024; Kim and Lee, 2023). Additionally, AI chatbots are being used more often to assist students with tasks such as technical and argumentative writing (El-Magd, 2022; Su, Lin, and Lai, 2023). Despite their widespread implementation, a systematic tool for evaluating AI chatbot's efficiency remains elusive (Jain et al., 2018; Maroengsit et al., 2019). However, the lack of a unified framework for evaluating these chatbots poses a significant barrier to their effective integration (Gregori-Giralt and Menéndez-Varela, 2019; Maroengsit et al., 2019). Andrade and Heritage (2017) and Brookhart (2019) assert that rubrics are crucial in setting clear expectations for learner competencies and within any educational tool's assessment framework. They serve as critical instruments in scaffolding both formative and summative assessment processes, enabling the tracking of progress and ensuring alignment with educational standards (Andrade, 2010; Darling-Hammond, Newton, and Wei, 2013).

Rooting its methodology in a rigorous validation process, it demonstrated the rubric's potential applicability and reliability in real-world educational settings and its capacity to standardize the evaluation of AI chatbots' effectiveness in aiding task-solving activities among students.

The intensive investigation into the reliability and authenticity of the proposed tool resulted in consistent scoring outcomes among faculty reviewers. These results are in line with Holsti's (1970) benchmark for instrument strength with an agreement rate of 85% or more. This consistency is pivotal for a precise and equitable evaluation of chatbot functionality. Factor analysis revealed two significant domains: "Quality of Content" and "Quality of Expression." This enhancement of the scale's construct validity established its convergent validity, highlighting the importance of content validity in rubric assessments (Gregori-Giralt and Menéndez-Varela, 2021). Statistical methods such as the Content Validity Ratio (CVR) and a modified kappa coefficient fortified the evaluation tool's reliability and validity (Davis, 1992; Gilbert and Prion, 2016). This suggests that a high Content Validity Index (CVI) value of 0.926 and acceptance of CVR-based criteria substantiate the scale's reliability and validity. A significant finding was the convergent validity, confirmed through factor analysis. The Kaiser-Meyer-Olkin test, as recommended by Hair et al. (2014), verified the factor analysis's effectiveness, strengthening the validity of the assessment scale. Furthermore, the researchers assessed the assessment scale's internal consistency reliability using the Pearson correlation and Cronbach's alpha coefficients, indicating its robustness.

These findings are in accordance with findings reported by Hill, Ford, and Farreras (2015). As an illustration, our investigation adds to the larger conversation on evaluating chatbot performance. The findings extend the discourse on rubric evaluations. The results confirm the reliability and validity of the assessment scale in assessing chatbot performance and open avenues for its use in enhancing chatbot designs and identifying areas for improvement. Here, we compared the results of the proposed method with those of the traditional methods.

These results go beyond previous reports. For instance, the validation of the rubric, underscored by the unanimous acceptance of criteria such as "Accuracy," "Relevance," and "Efficiency," echoes the critical attributes highlighted in the literature for evaluating educational tools (Almasre, 2024; Cope, Kalantzis, and Searsmith, 2021). These attributes are important to ensure that AI chatbots effectively aid pedagogy, fostering both knowledge acquisition and the development of critical thinking skills among students. The rigorous statistical validation, including the high agreement percentages among assessors and the substantial reliability coefficients, attests to the rubric's robustness, aligning with best practices in educational assessment (Bradley, Anderson, and Eagle, 2020). Furthermore, the factor analysis revealing two distinct domains - "Quality of Content" and "Quality of Expression" - validates the conceptual framework proposed by this study and offers critical insights into the multifaceted nature of evaluating AI chatbots. This distinction underscores the complexity of assessing AI chatbots, where factual accuracy and communicative effectiveness are paramount. This detailed approach to evaluation aligns with the Competency-Based Learning (CBL) framework, which prioritizes mastery of essential competencies and underscores the significance of feedback in learning processes (Brown, 2012; Henri, Johnson, and Nepal, 2017). Additionally, the validation process of the rubric revealed its robustness in assessing the intended competencies, reinforcing the importance of a balanced approach to AI integration in educational contexts. Abbas, Jam, and Khan (2024) noted that this balance is crucial to harness AI's innovative capabilities while mitigating potential negative impacts on student learning behaviors and outcomes.

This result ties well with previous studies wherein the current study's findings resonate with and diverge from the literature in several key areas. Unlike previous research by García-Orosa, Canavilhas and Vázquez-Herrero (2023) and Yang (2022), which emphasized the potential and challenges of integrating AI in education without a clear framework for evaluation, this study provides a concrete rubric for assessing the effectiveness of AI chatbots in educational settings. The rubric's focus on "Accuracy," "Relevance," and "Efficiency" parallels the attributes identified by Almasre (2024) and Cope, Kalantzis, and Searsmith (2021) as essential for educational tools. However, our findings extend beyond these attributes by validating a comprehensive set of criteria through empirical methods, addressing a gap in the literature regarding the systematic assessment of AI chatbots. The distinction between "Quality of Content" and "Quality of Expression" identified through factor analysis further deepens the understanding of chatbot assessment. This approach offers a more comprehensive framework than the general discussions on AI chatbot capabilities presented by Kim and Lee (2023) and Hmoud et al. (2024). They emphasized the benefits of AI chatbots in fostering creativity and engagement without specifying mechanisms for evaluation.

Rubrics, as elucidated by Andrade and Heritage (2017) and Brookhart (2019), serve as essential tools in teacher education by making explicit the competencies expected of learners. They facilitate both formative and summative assessments, clearly communicating expectations and tracking progress over time (Darling-Hammond, Newton, and Wei, 2013). Based on this foundational understanding, our study expands the application of rubrics to AI chatbots, portraying these technological tools as 'learners' whose performance and integration into educational practices require careful evaluation.

The study identifies six items: "Accuracy," "Relevance," "Efficiency," "Coherence," "Comprehensiveness," "Grammar and Spelling," "Argument and Evidence," "Language and Tone," and "Creativity and Originality" as key criteria. This selection highlights the importance of content quality and communicative effectiveness in educational AI chatbots, offering a clear response to the needs within educational settings for reliable and engaging AI tools. Applying the Task-Solving Assessment Rubric demonstrates that AI chatbots can significantly support task-solving assessments when evaluated against the identified criteria. They offer a means to enhance learning engagement and creativity and ensure that students interact with AI technologies that meet high accuracy, relevance, and efficiency standards. This finding validates the hypothesis that properly assessed and integrated AI chatbots are valuable in higher education.

Our study's outcomes significantly contribute to the discourse on AI chatbot evaluation in education. We bridge the identified literature gap with a validated assessment tool reflecting the core principles of CBL and Brown's developmental stages (Henri, Johnson, and Nepal, 2017; Tenakwah et al., 2023). While this research marks a critical step toward systematic AI chatbot assessment, it also highlights the necessity for ongoing validation efforts. Future studies should extend this work across varied educational contexts, chatbot types, and learning tasks while also considering the ethical implications of AI in education (Korteling et al., 2021; Lim, 2022; Tate et al., 2023).

5. Conclusion, Recommendations, Limitations, and Implications for Future Research

This research pioneered the developing and validation of a rubric-based assessment scale for evaluating AI chatbot performance in educational settings. By employing rigorous methodology and a thorough validation process, the study has established a foundation for systematically evaluating chatbot effectiveness, filling a significant gap in the literature identified by Jain et al. (2018) and Maroengsit et al. (2019). Insights from this study suggest various directions for future research and practical application.

To enhance the assessment tool's generalizability and applicability across various educational contexts, future research should aim to validate the rubric with a broader, more diverse sample (Nsabayezu et al., 2022). Investigating the assessment scale's effectiveness in evaluating chatbot performance over time could yield insights into the durability and evolution of chatbot effectiveness. There is a pressing need to examine the ethical implications of AI chatbot use, such as potential overreliance on AI, the quality of AI-generated writing, and issues related to literacy assessment (Korteling et al., 2021; Lim, 2022; Tate et al., 2023).

Enhancing the rubric-based assessment with user feedback and performance metrics would give a broader understanding of chatbot effectiveness. This combines users' feelings with performance data (Jain et al., 2018). Furthermore, future research could specifically target the relationship between chatbot performance and learning outcomes, such as knowledge retention, critical thinking, and problem-solving abilities (Liu, 2017).

There are several limitations to this research. First, The sample size and diversity were constrained, potentially impacting the rubric's generalizability across different educational settings and subjects. Additionally, the study focused on immediate assessment outcomes, leaving room for exploration of long-term impacts and the sustainability of chatbot effectiveness over time. The implications of this research are manifold. The validated rubric offers educators and technologists a practical tool for assessing and improving AI chatbot integration in educational contexts. It underscores the necessity of aligning AI technologies with pedagogical objectives and competency-based learning frameworks, a critical insight that aligns with the foundational principles discussed by Akgun and Greenhow (2021). For the broader field of educational technology, this study highlights the importance of developing reliable, validated tools for evaluating emerging technologies. It highlights how AI chatbots can improve educational experiences when assessed and implemented well. This echoes what Korteling et al. (2021), Lim (2022), and Tate et al. (2023) have said about the importance of considering AI's ethical and practical implications in education. In conclusion, this research contributes significantly to the dialogue on AI chatbot performance assessment in education, presenting a validated assessment tool and outlining a path for future research. As we advance, we must continue to explore, validate, and refine our methods, ensuring that AI chatbots and similar technologies are leveraged to their fullest potential in enriching the educational landscape.

References

Abbas, M., Jam, F.A. and Khan, T., 2024. Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, [online] 21(1). https://doi.org/10.1186/s41239-024-00444-7

Abdul-Kader, S.A. and Woods, J., 2015. Survey on Chatbot Design Techniques in Speech Conversation Systems. *International Journal of Advanced Computer Science and Applications*, [online] 6(7). https://doi.org/10.14569/ijacsa.2015.060712

Akgün, S. and Greenhow, C., 2022. Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI And Ethics*, [online] 2(3), pp.431–440. https://doi.org/10.1007/s43681-021-00096-7

Almasre, M.A., 2024. Development and evaluation of a custom GPT for the assessment of students' designs in a typography course. *Education Sciences*, [online] 14(2), p.148. https://doi.org/10.3390/educsci14020148

Ayre, C. and Scally, A.J., 2014. Critical values for LawsHe's Content Validity ratio. *Measurement and Evaluation in Counseling and Development*, [online] 47(1), pp.79–86. https://doi.org/10.1177/0748175613513808

Baidoo-Anu, D. and Ansah, L.O., 2023. Education in the Era of Generative Artificial Intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, [online] 7(1), pp.52–62. https://doi.org/10.61969/jai.1337500

Bradley, E.J., Anderson, S.E. and Eagle, L., 2020. Use of a marking rubric and self-assessment to provide feedforward to level 5 undergraduate Sport students: student perceptions, performance and marking efficiency. *Journal of Learning Development in Higher Education*, [online] (18). https://doi.org/10.47408/jldhe.vi18.557

Brandtzæg, P.B. and Følstad, A., 2017. Why people use chatbots. In: *Lecture notes in computer science*. [online] pp.377–392. https://doi.org/10.1007/978-3-319-70284-1_30

Brookhart S.M., 2013. *How to create and use rubrics for formative assessment and grading*. [online] CiNii Books. Available at: <http://ci.nii.ac.jp/ncid/BB18336410>.

Brookhart, S.M., 2018. Appropriate criteria: key to effective rubrics. *Frontiers in Education*, [online] 3. https://doi.org/10.3389/feduc.2018.00022

Brown J.D., 2012. *Developing, using, and analyzing rubrics in language assessment with case studies in Asian and Pacific languages*. [online] CiNii Books. Available at: <http://ci.nii.ac.jp/ncid/BB17593429>.

Chi, Y., 2013. Developing, using, and analyzing rubrics in language assessment with case studies in Asian and Pacific languages. Edited by J. D. Brown. *Language Assessment Quarterly*, [online] 10(2), pp.236–239. https://doi.org/10.1080/15434303.2013.769553

Cope, B., Kalantzis, M. and Searsmith, D., 2020. Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational Philosophy and Theory (Print)*, [online] 53(12), pp.1229–1245. https://doi.org/10.1080/00131857.2020.1728732

Davis, L.L., 1992. Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, [online] 5(4), pp.194–197. https://doi.org/10.1016/s0897-1897(05)80008-4

De La Rosa Gómez, A., Cano, J.M.M. and Díaz, G.A.M., 2019. Validation of a rubric to evaluate open educational resources for learning. *Behavioral Sciences*, [online] 9(12), p.126. https://doi.org/10.3390/bs9120126

De Vera, R.L., 2023. Formative Assessment and Metacognition towards Relevant Learning. *European Journal of Science, Innovation and Technology*, [online] 3(1), pp.54–65. Available at: <https://ejsit-journal.com/index.php/ejsit/article/view/158>.

El-Magd, M.A.A.E.-M., 2022. Text Chatbot Assisted Edublogs for Enhancing the EFL Technical Writing Performance among Computer and Informatics Students. *MaǧAllaẗ Kulliyyaẗ Al-Tarbiyyaẗ Fi Al-ʿulūM Al-Tarbawiyyaẗ*, [online] 46(2), pp.99–145. https://doi.org/10.21608/jfees.2022.241755

Field, A.P., 2018. *Discovering statistics using IBM SPSS statistics*. [online] Available at: <https://dl.acm.org/citation.cfm?id=2502692>.

Flores-Vivar, J.M.F. and García-Peñalvo, F.J., 2023. Reflections on the ethics, potential, and challenges of artificial intelligence in the framework of quality education (SDG4). *Comunicar*, [online] 31(74), pp.37–47. https://doi.org/10.3916/c74-2023-03

García-Orosa, B.G., Canavilhas, J. and Vázquez-Herrero, J., 2023. Algorithms and communication: A systematized literature review. *Comunicar*, [online] 31(74), pp.9–21. https://doi.org/10.3916/c74-2023-01

Gilbert, G.E. and Prion, S., 2016. Making sense of methods and measurement: Lawshe's Content Validity Index. *Clinical Simulation in Nursing*, [online] 12(12), pp.530–531. https://doi.org/10.1016/j.ecns.2016.08.002

Gorsuch, R.L., 2014. *Factor Analysis: Classic Edition*. [online] Available at: <https://www.amazon.com/Factor-Analysis-Psychology-Routledge-Editions/dp/1138831999>.

Gregori-Giralt, E. and Menéndez-Varela, J.-L., 2021. The content aspect of validity in a rubric-based assessment system for course syllabuses. *Studies in Educational Evaluation*, [online] 68, p.100971. https://doi.org/10.1016/j.stueduc.2020.100971

Hack, C., 2015. Analytical rubrics in higher education: A repository of empirical data. *British Journal of Educational Technology*, [online] 46(5), pp.924–927. https://doi.org/10.1111/bjet.12304

Henri, M., Johnson, M.D. and Nepal, B., 2017b. A review of Competency-Based Learning: Tools, assessments, and recommendations. *Journal of Engineering Education*, [online] 106(4), pp.607–638. https://doi.org/10.1002/jee.20180

Hill, J., Ford, W.R. and Farreras, I.G., 2015. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, [online] 49, pp.245–250. https://doi.org/10.1016/j.chb.2015.02.026

Hmoud, M. and Shaqour, A., 2024. AIEd Bloom's Taxonomy: A Proposed Model for Enhancing Educational Efficiency and Effectiveness in the Artificial Intelligence Era, *the International Journal of Technologies in Learning*, 31(2), pp. 111–128. https://doi.org/10.18848/2327-0144/cgp/v31i02/111-128

Hmoud, M., Swaity, H., Hamad, N., Karram, O. and Daher, W., 2024. Higher education students' task motivation in the Generative Artificial Intelligence context: the case of ChatGPT. *Information*, [online] 15(1), p.33. https://doi.org/10.3390/info15010033

Holsti, O.R., 1970. Content analysis for the Social Sciences and Humanities. *American Sociological Review*, [online] 35(2), p.356. https://doi.org/10.2307/2093233

Ilieva, G., Yankova, T., Klisarova-Belcheva, S., Dimitrov, A., Bratkov, M. and Angelov, D., 2023. Effects of generative chatbots in higher education. *Information*, [online] 14(9), p.492. https://doi.org/10.3390/info14090492

Jain, M., Kumar, P., Kota, R. and Patel, S.N. 2018. Evaluating and Informing the Design of Chatbots. In Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18). Association for Computing Machinery, New York, NY, USA, [online] pp. 895–906. https://doi.org/10.1145/3196709.3196735

Jönsson, A. and Svingby, G., 2007. The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review (Print)*, [online] 2(2), pp.130–144. https://doi.org/10.1016/j.edurev.2007.05.002

Kim, J. and Lee, S.-S., 2022. Are two heads better than one?: The Effect of Student-AI Collaboration on Students' Learning Task Performance. *TechTrends (Online)*, [online] 67(2), pp.365–375. https://doi.org/10.1007/s11528-022-00788-9

Kooli, C., 2023. Chatbots in Education and Research: A Critical Examination of ethical implications and solutions. *Sustainability (Basel)*, [online] 15(7), p.5614. https://doi.org/10.3390/su15075614

Korteling, J.E., Van De Boer-Visschedijk, G.C., Blankendaal, R., Boonekamp, R. and Eikelboom, A.R., 2021. Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence (Lausanne)*, [online] 4. https://doi.org/10.3389/frai.2021.622364

Lameras, P. and Arnab, S., 2021. Power to the Teachers: An Exploratory review on Artificial intelligence in education. *Information (Basel)*, [online] 13(1), p.14. https://doi.org/10.3390/info13010014

Lim, V. F. 2022. ChatGPT raises uncomfortable questions about teaching and classroom learning, *The Straits Times*, Available at: <https://www.straitstimes.com/opinion/needto-review-literacy-assessment-in-the-age-of-chatgpt>.

Linardatos, P., Papastefanopoulos, V. and Kotsiantis, S., 2020. Explainable AI: A review of Machine Learning Interpretability Methods. *Entropy (Basel. Online)*, [online] 23(1), p.18. https://doi.org/10.3390/e23010018

Liu, S. H. 2017. Trends of digital learning in K-12 schools: A review of the literature on research and practice, *Journal of Educational Technology Development and Exchange*, 10(2), pp. 23-42.

Luger, E. and Sellen, A. 2016. Like having a really bad PA: The gulf between user expectation and experience of conversational agents, in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM.

Maroengsit, W. Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., and Theeramunkong, T. 2019. A survey on evaluation methods for chatbots, in Proceedings of the 2019 7th International Conference on Information and Education Technology. New York, NY, USA: ACM, pp. 111–119. Available at: <https://doi.org/10.1145/3323771.3323824>.

McMurtrie, B. 2023. Teaching: Will ChatGPT change the way you teach? *The Chronicle of Higher Education*. Available at: <https://www.chronicle.com/newsletter/teaching/2023-01-05>.

Meo, S.A., Al-Masri, A.A., Alotaibi, M., Meo, M.Z.S. and Meo, M.O.S., 2023. CHATGPT Knowledge Evaluation in Basic and Clinical Medical Sciences: Multiple Choice Question Examination-Based Performance. *Healthcare (Basel)*, [online] 11(14), p.2046. https://doi.org/10.3390/healthcare11142046

Mishra, S., 2017. Open educational resources: removing barriers from within. *Distance Education*, [online] 38(3), pp.369–380. https://doi.org/10.1080/01587919.2017.1369350

Moore, K.B., Bonnett, R. and Colbert-Getz, J.M., 2020. A process and rubric for a group to review the quality of a Medical Education Course/Clerkship. *MedEdPORTAL*. [online] https://doi.org/10.15766/mep_2374-8265.10911.

Nsabayezu, E., Mukiza, J., Iyamuremye, A., Mukamanzi, O.U. and Mbonyiryivuze, A., 2022. Rubric-based formative assessment to support students' learning of organic chemistry in the selected secondary schools in Rwanda: A technology-based learning. *Education and Information Technologies*, [online] 27(9), pp.12251–12271. https://doi.org/10.1007/s10639-022-11113-5

Panadero, E. and Jönsson, A., 2013. The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, [online] 9, pp.129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Reddy, Y.M. and Andrade, H.L., 2010. A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, [online] 35(4), pp.435–448. https://doi.org/10.1080/02602930902862859

Sedoc, J., Ippolito, D., Kirubarajan, A., Thirani, J., Ungar, L.H. and Callison-Burch, C., 2018. ChatEval: a tool for the systematic evaluation of chatbots. *Association for Computational Linguistics*, [online] pp.42–44. https://doi.org/10.18653/v1/w18-6709

Stanley, T. 2021. What is a rubric?, in *Using RUBRICS for Performance-Based Assessment*. New York: Routledge, pp. 9–20.

Stanley, T., 2021. How to write a rubric. In: *Routledge eBooks*. [online] pp.53–68. https://doi.org/10.4324/9781003239390-6

Stiggins, R.J. 1997. Student-centered classroom assessment, *New York: Merrill*. Available at: <https://files.nwesd.org/depts/eadmin/Admin_Website/CIT-CL/LiteratureReference/JournalArticles/Student-Centered-Classroom-Assessment_Stiggins.pdf>.

Streiner, D.L., 2003. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, [online] 80(1), pp.99–103. https://doi.org/10.1207/s15327752jpa8001_18

Su, Y., Lin, Y.G. and Lai, C., 2023. Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, [online] 57, p.100752. https://doi.org/10.1016/j.asw.2023.100752

Tan, K., 2020. *Assessment rubrics decoded*. [online] *Routledge eBooks*. https://doi.org/10.4324/9780429022081

Tate, T., Doroudi, S., Ritchie, D., Xu, Y. and Warschauer, M., 2023. Educational Research and AI-Generated Writing: Confronting the Coming Tsunami. *EdArXiv*. [online] https://doi.org/10.35542/osf.io/4mec3

Tenakwah, E.S., Boadu, G., Tenakwah, E.J., Parzakonis, M., Brady, M., Kansiime, P., Said, S., Ayilu, R.K., Radavoi, C.N. and Berman, A.L., 2023. Generative AI and Higher Education Assessments: A Competency-Based Analysis. *Research Square (Research Square)*. [online] https://doi.org/10.21203/rs.3.rs-2968456/v2

Tilden, V.P., Nelson, C.A. and May, B.A., 1990. Use of qualitative methods to enhance content validity. *Nursing Research*, [online] 39(3), p.172???175. https://doi.org/10.1097/00006199-199005000-00015

Turney, S., 2022. *Pearson Correlation Coefficient (r) | Guide & Examples*. [online] Scribbr. Available at: <https://www.scribbr.com/statistics/pearson-correlation-coefficient>.

Vicente-Yagüe-Jara, M.-I., Martínez, O.L., Navarro-Navarro, V. and Cuéllar-Santiago, F., 2023. Writing, creativity, and artificial intelligence. ChatGPT in the university context. *Comunicar*, [online] 31(77). https://doi.org/10.3916/c77-2023-04

Wilson, F.G., Pan, W. and Schumsky, D.A., 2012. Recalculation of the critical values for LawsHE's content validity ratio. *Measurement and Evaluation in Counseling and Development*, [online] 45(3), pp.197–210. https://doi.org/10.1177/0748175612440286

Yang, W., 2022. Artificial Intelligence education for young children: Why, what, and how in curriculum design and implementation. *Computers & Education: Artificial Intelligence*, [online] 3, p.100061. https://doi.org/10.1016/j.caeai.2022.100061

Generative-AI, a Learning Assistant? Factors Influencing Higher-Ed Students' Technology Acceptance

Kraisila Kanont[1], Pawarit Pingmuang[1], Thewawuth Simasathien[1], Suchaya Wisnuwong[1], Benz Wiwatsiripong[1], Kanitta Poonpirome[1], Noawanit Songkram[3] and Jintavee Khlaisang[2]
[1]Department of Educational Technology and Communications, Faculty of Education, Chulalongkorn University, Thailand
[2]Center of Excellence in Educational Invention and Innovation, Department of Educational Technology and Communications, Faculty of Education, Chulalongkorn University, Thailand
[3]Learning Innovation for Thai Society Research Unit (LIFTS), Chulalongkorn University, Bangkok, Thailand

kraisila.kan@kmutt.ac.th
pawarit.p@me.com (Correspondence author)
t008@mbp.ac.th
sunachayaw@gmail.com
Bybenz@gmail.com
Kanitta.poonpirome@gmail.com
noawanit.s@chula.ac.th
jintavee.m@chula.ac.th

https://doi.org/10.34190/ejel.22.6.3196

Abstract: This study investigates the factors influencing the adoption of Generative-AI tools amongst Thai university students, employing the Technology Acceptance Model (TAM) as a theoretical framework. Data from 911 higher education students from 10 different Thai Universities Health Sciences, Sciences and Technology, Social Sciences and Humanities, and Vocational Fields were analysed via Structural Equation Modelling (SEM). The instrument used in collecting the data was a questionnaire. Results indicated that Expected Benefits, Perceived Usefulness, Attitude Toward Technology, and Behavioural Intention all significantly impacted student adoption of Generative AI. Intriguingly, Perceived Ease of Use was negatively correlated with Perceived Usefulness, challenging conventional TAM assumptions. This study underscores the need to address language barriers, foster a culture of innovation, and establish ethical guidelines to promote responsible AI use within education. Despite inherent limitations, this research contributes to our understanding of AI adoption in educational settings and helps inform strategies for equitable access and responsible innovation. The result demonstrated that the easier a tool was to use, the less value leaners seemed to see in it for their learning process. It can be implied that as Generative-AI get more intuitive, learners think they're less helpful. These finding challenges a few of those assumptions we usually make within the TAM model. It also points out the characteristic of learners which affects their learning preferences and expectation. Another finding showed the impact of language barrier on non-native English speaker that obstruct the user experience in AI services. Moreover, the role of universities in fostering both AI integration for learning for and the ethical implementation of Generative AI. By providing a supportive environment that encourages AI experimentation, redesign learning, empowering learners and faculty instructors to investigate how Generative AI can be applied across disciplines, and developing guidelines for ethical use, universities play a critical role in shaping the effective and responsible integration of AI into the next educational landscape.

Keywords: Artificial Intelligence in education, Educational technology, Generative-AI, Student perceptions, Technology Acceptance Model, SEM research

1. Introduction

Education is an ever-evolving field significantly influenced by technological advancements. One of the pivotal drivers of change in educational development has consistently been technology (Dwivedi et al., 2021; Murugesan and Cherukuri, 2023). In this continuum of technological evolution, Artificial Intelligence (AI) has emerged as a transformative force, reshaping learning experiences and pedagogical approaches.

Recent advancements have seen the rise of Generative-AI, a subset of AI that autonomously creates content and data, distinguishing itself from other AI forms primarily focusing on data analysis or interpretation. This type of AI has the unique capability to generate new, personalised content, thereby offering significant potential to

Cite this article: Kanont, K. et al. 2024. "Generative-AI, a Learning Assistant? Factors Influencing Higher-Ed Students' Technology Acceptance", *Electronic Journal of e-Learning*, 22(6), pp 18-33, https://doi.org/10.34190/ejel.22.6.3196

revolutionise educational methodologies and learning experiences (Cooper, 2023; Dai, Liu and Lim, 2023; Zawacki-Richter et al., 2019).

However, integrating Generative-AI into the education sector is not without challenges. These include the need for a comprehensive approach to ensure sustainable development, inclusivity, and equity in AI applications within education (Ahmad et al., 2021; Dai, Liu and Lim, 2023; UNESCO, 2023a, UNESCO 2023b). Additionally, there is a pressing need for inclusive data systems and adequate preparation of educators and students for an AI-enhanced educational landscape (Hutson et al., 2022; Rasul et al., 2023).

The potential of Generative-AI in education is particularly notable in creating personalised and unique educational materials. It can understand and adapt to individual students' learning patterns and needs, offering customised learning experiences that can significantly enhance engagement and learning outcomes (Celik et al., 2022; Gimpel et al., 2023). Such personalised approaches can lead to more effective learning, catering to individual preferences and learning styles, thereby potentially improving the overall quality of education.

Nevertheless, the adoption of Generative AI in higher education, especially in Thailand, is still nascent. There are various challenges to its widespread adoption, including concerns over its ability to understand and interpret complex educational content, ethical considerations, issues of plagiarism, and maintaining academic integrity (Chan and Hu, 2023; Nguyen, 2023; Wang et al., 2023; Su and Yang, 2023).

Addressing these challenges, this study explores the factors influencing the adoption of Generative-AI technologies in Thai higher education. It seeks to understand how Thai higher education students perceive Generative-AI and identify the key factors influencing its acceptance and integration into educational practices. By examining these factors, the study intends to contribute valuable insights toward developing effective strategies that enhance learning outcomes, ensure equitable access to AI's benefits, and maintain academic integrity in an increasingly AI-integrated educational landscape (Song, 2024).

This paper presents the factors influencing the Technology Acceptance Model (TAM) amongst Thai university students, by Structural Equation Modelling to clarify the relationships that reflect the actual use of Generative-AI. This study reviews related literature to identify the factors in the development of a research framework and hypotheses between TAM and Generative-AI. The outlines are included the research methodology, sample sampling procedures, instruments, and data analysis. The results, discussion, and implications are subsequently presented, emphasising the key findings on the relationships between TAM factors. This paper concludes with recommendations for the application of Generative-AI in higher education, as well as suggestions for future research and a discussion of the study’s limitations.

1.1 Research Questions

RQ1: How do higher education students perceive the Expected Benefit, Perceived Usefulness, and Perceived Ease of Use of Generative-AI, and how do these perceptions influence their Attitude Toward Using and Behavioural Intention to use Generative-AI, as well as its Actual Use?

RQ2: What are the relationships between Expected Benefit, Perceived Usefulness, and Perceived Ease of Use of Generative-AI, and Behavioural Intention, and how do these relationships affect the Actual Use of Generative-AI in higher education?

RQ3: How do Expected Benefits, Perceived Usefulness, Perceived Ease of Use, and Behavioural Intention influence the adoption and integration of Generative-AI technologies in higher education?

1.2 Research Objectives

This research aims to investigate the factors influencing the adoption of Generative-AI amongst higher education students using the TAM. The specific objectives are:

- To study the Expected Benefit, Perceived Usefulness, Perceived Ease of Use, Attitude Toward Using Generative-AI, Behavioural Intention, and Actual Use of Generative-AI amongst higher education students.
- To study the relationship between Expected Benefit, Perceived Usefulness, Perceived Ease of Use, Attitude toward Generative-AI, Behavioural Intention, and Actual Use of Generative-AI amongst higher education students.
- To determine the influences of Expected Benefit, Perceived Usefulness, Perceived Ease of Use, and Behavioural Intention towards Generative-AI adoption.

2. Literature Review and Hypothesis Development

2.1 Generative-AI in Higher Education

Generative-AI has been increasingly integrated into the educational sector, providing a transformative approach to teaching and learning (Lim et al., 2023). It has been utilised to create personalised learning experiences, enhance student engagement, and improve educational outcomes (Gustafson, 2023). Generative-AI tools, such as chatbots and virtual tutors, have facilitated interactive learning environments, provided instant feedback, and supported personalised learning paths (Gustafson, 2023). Generative-AI can be divided into four main types: text generation, image creation, and video production (Jiayang Wu et al., 2023).

- *Text generation*: Generative-AI includes the areas of structured composition, imaginative writing, and conversational scripting as its primary branches such as ChatGPT, Google Gamini, Claude, etc.
- *Image generation*: Utilizing Generative-AI allows individuals to modify and introduce additional components into their images in response to specific instructions such as DALL-E, Midjourney, etc.
- *Sound generation*: Generative-AI in audio involves two main types: synthesis of speech from text and replicating existing voices such as MURF, Soundraw, Botnoi, etc.
- *Video generation*: The application of Generative-AI in creating video content is employed in making movie trailers and advertising clips such as Synthesia, Maverick, etc.

Owing to the diversity of Generative-AI, which can serve as a learning aid amongst students, adopting AI in higher education is a topic of increasing interest and relevance. AI technologies, including Generative-AI, have the potential to revolutionise the way education is delivered, making it more personalised, efficient, and effective for learning in higher education (Lim et al., 2023; Sandu and Gide, 2019). Some studies adopted Generative-AI as a tool with constructivism learning theory, such as knowledge co-creation and collaborative learning (Salinas-Navarro et al, 2024, Zhou and Schofield, 2024)

One of the key factors influencing the adoption of AI in higher education is the technology's perceived usefulness and ease of use. This is consistent with the Technology Acceptance Model (TAM), which posits that these two factors significantly influence the intention to use technology (Davis, 1989). In the context of AI in education, perceived usefulness could be related to the potential of AI to enhance teaching and learning, while ease of use could be associated with the user-friendliness of the AI system (Lim et al., 2023).

Therefore, the integration of Generative-AI entails a process of endorsement through which an individual's behaviour demonstrates acceptance and engagement. The utilisation and cognitive development behaviours associated with Artificial Intelligence as an educational instrument enhance the efficacy of learning, facilitating the production of informational and media data, thereby improving educational outcomes (Kelly et al., 2023; Li, 2023; Pillai et al., 2023).

2.2 Technology Acceptance Model

To explore the receptivity and application of Generative-AI technologies amongst Thai higher education students, this investigation employs the Technology Acceptance Model (TAM) proposed by Davis (1989). This model provides a comprehensive theoretical framework for understanding the adoption and acceptance of technologies within information systems domains. In several years, the TAM has developed into an authoritative model elucidating the acceptance, refusal, and use of innovative technologies across diverse disciplines, including digital learning, information and communication technologies, and educational technology (Khlaisang, Teo, and Huang, 2019, Liu G et al., 2022, Ma and Huo, 2023)

The Technology Acceptance Model (TAM), as proposed by Davis (1989), suggests that users' acceptance and adoption of technology can be anticipated through their Behavioural Intention (BI) to use it. Behavioural intention refers to a person's belief in an action or behaviour that is about to happen in the future by predicting the outcome or impact of that action). BI can be measured using three types of questions: expect, want, and intend (Chuenphitthayavut, Zihuang and Zhu, 2020). BI can positively or negatively impact yourself or others, provided sufficient resources about user attitudes and subjective norms are provided for that action. Attitude has a significant positive effect on the intention to use AI. Personal concerns significantly negatively affect the intention to use AI (Cao et al., 2021).

In the TAM framework, BI is influenced by users' Perceived Usefulness (PU) of the technology in accomplishing specific tasks and the Perceived ease of use (PEOU) with which they can employ the technology. PU refers to an individual's awareness that technology can help improve learning performance. The awareness of its usefulness influences the choice and use of technology in learning. Including attitudes that affect acceptance and demonstrate a willingness to increase learning efficiency (Dhingra and Mudgal, 2019; Nugroho, Dewanti and Novitasari, 2018). Then, PEOU means "the degree to which a person believes using a particular system will be effortless." Rosenberg (1983) stated that many people's acceptance of technology depends on their use and learning to use it themselves. Learning how to use technology for people in society will result in continuously developing and improving technology. Sallam et al. (2023) state that perceived ease of use is the user's expectation that technology can be used quickly and effortlessly. The technology must be easily recognisable and have no complexity.

Moreover, PU and PEOU are influenced by Expected Benefits (EB), which means the degree to which students believe that AI applications improve the quality of learning and education. AI provides many ready-made programmes for self-learning or teacher-assisted learning (Al Darayseh, 2023; Nazaretsky, Cukurova, and Alexandron, 2022).

The model's attitudinal variables address Attitudes toward using (AT) or their BI towards technology use, typically quantified through the Intention to Use as a marker of attitudinal readiness towards embracing specific technologies (Davis, 1989; Venkatesh, 2000). According to Kim et al. (2020), technology is believed to be beneficial for users and easy to use, with a cheerful outlook towards its usage. In this study, the attitude toward using artificial intelligence consists of the perception that AI is easy to use and valuable. The positive or negative feelings of an individual towards the use of cognitive artificial intelligence in the educational process (Al-Adwan et al., 2023; Chatterjee and Bhattacharjee, 2020; Chou et al., 2022; Cruz-Benito et al., 2019; Sing et al., 2022).

Then, the foundational TAM elements in this research consist of EB, PEOU, PU, AT, BI, and AU. These components underscore a progression from cognitive recognition through attitudinal response to Actual use (AU) (Al-Emran, Mezhuyev, and Kamaludin, 2018; Davis, 1989). Furthermore, TAM encourages the exploration of AU patterns, thereby connecting theoretical constructs with practical observations. As Davis (1989) elucidates, the perceived user-friendliness of an application significantly enhances its perceived utility, thereby increasing the likelihood of its acceptance and adoption by individuals.

This study aims to investigate the application of Generative-AI technologies by Thai higher education students. The conceptualisations of the primary variables within the Technology Acceptance Model (TAM) have been refined and adopted. Hence, we set the operational definition in TAM for this study as follows in the Table 1.

Table 1: Definitions of main TAM components in this study

TAM components	**Definitions**
Expected Benefits (EB)	The degree to which students anticipate significant improvements in their learning outcomes because of integrating Generative-AI technologies into their learning.
Perceived Usefulness (PU)	The degree to which students are convinced that using Generative-AI technologies will significantly contribute to their learning.
Perceived Ease of Use (PEOU)	The degree to which students perceive that employing Generative-AI technologies will necessitate minimal exertion.
Attitude Toward Using (AT)	The degree of students' favourable or unfavourable evaluation regarding adopting Generative-AI technologies in their learning process.
Behavioural Intention (BI)	The degree of preparedness amongst learners to incorporate Generative-AI technologies into their educational endeavours.
Actual Use (AU)	The degree to which students effectively employ Generative-AI technologies within their educational context.

Hence, to clarify the hypothesised relationships between factors in the quantitative component of our study, we have introduced a structural model depicted in Figure 1. This model is grounded in the Technology Acceptance Model (TAM) and research hypotheses:

H1: EB positively influences PU of Generative-AI amongst higher education students.

H2: EB positively influences PEOU of Generative-AI amongst higher education students.

H3: PEOU positively influences PU of Generative-AI amongst higher education students.

H4: PEOU positively influences AT towards Generative-AI amongst higher education students.

H5: PU positively influences AT towards Generative-AI amongst higher education students.

H6: AT positively influences BI to use Generative-AI amongst higher education students.

H7: BI positively influences the AU of Generative-AI amongst higher education students.

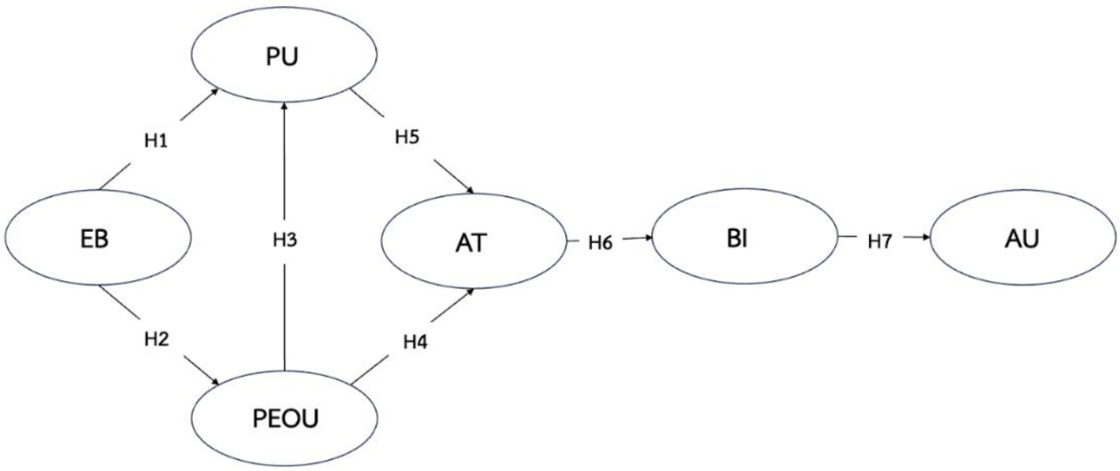


Figure 1: Research model of study

3. Research Methodology

3.1 Samples

In this research, the sample consisted of 911 students from 10 different Thai Universities Health Sciences, Sciences and Technology, Social Sciences and Humanities, and Vocational Fields, diversified in terms of academic level, gender, and field of study. The convenience random sampling method was used to distribute research instrument to samples for this study in each Thai Universities. Hair et al. (2018) suggested that the sample size was 10–20 samples per item for applying structural equation modelling (SEM). The current sample size of 911 with 6 constructs of 31 items was also considered fit and above (911 > 31*20 = 620) the desired level. Therefore, the sample size was considered appropriate to conduct SEM.

The researchers followed the privacy, data protection, and confidentiality policies. Therefore, all research participants' identities and data remained unidentifiable and will be deleted immediately after the research is completed. The data were consented by all participants for research purpose.

3.2 The Instrument

The research instrument, the survey questionnaire, was designed to gather data for the study and was divided into three parts. Part 1 focused on collecting demographic information from the sample, including gender, age, education levels, department, and year of study. This information is essential for understanding the characteristics and composition of the participants. Part 2 of the questionnaire aimed to gather information about the student's daily use and access to Generative-AI. This section explored the frequency and extent students engage with Generative-AI in their educational activities. Understanding students' current usage patterns provides insights into their familiarity with and exposure to Generative-AI. Part 3 of the questionnaire explored the students' perceptions and acceptance of Generative-AI. This section consisted of 31 items that assessed various dimensions of Technology Acceptance. Specifically, there were 5 items for EB, 6 items for PU, 5 items for PEOU, 5 items for AT, 5 items for BI, and 5 items for AU. Participants were asked to rate their agreement with each item on a 7-Likert scale ranging from "totally disagree" (1) to " totally agree" (7).

The evaluation of the quality of the Generative-AI Technology Acceptance Model (TAM) tool begins with establishing content validity through an extensive literature review. The researchers invited five subject matter experts in educational technology and artificial intelligence applications to assess the content validity of the measurement items about the study objectives. Regarding construct validity assessment, Confirmatory Factor Analysis (CFA) was conducted to examine the data's alignment with the hypothesized measurement model and the distinctiveness of the constructs. Adequate factor loadings and satisfactory fit indices indicate strong construct validity. Reliability analysis was performed by calculating Cronbach's alpha (CA) coefficients for each construct, which helped verify the internal consistency of the items, with values above 0.7 considered acceptable (Collier, 2020). The CA as internal consistency reliability values obtained from the data analysis for each construct were as follows: Expected Benefits = 0.88, Perceived Usefulness = 0.86, Perceived Ease of Use = 0.84, Attitude Toward Using = 0.81, Behavioural Intention = 0.86, and Actual Use = 0.78. These values were acceptable, indicating the reliability of the measurement items. Consequently, the researchers considered revising the wording of these items to enhance clarity and further distinguish them from other constructs.

Then, the researchers conducted a pilot test of the measurement instrument with a sample of 30 students to verify the clarity of the items and make appropriate adjustments to align with the context of higher education students in Thailand. This pilot testing with a representative sample of the target population enabled identifying and rectifying issues related to item clarity, response patterns, or scale reliability and validity. This iterative refinement process was crucial in enhancing the quality and applicability of the tool in assessing the acceptance of Generative-AI technologies, ensuring it effectively captures the nuances of this technology. The researcher initiated the research in March 2023. Following this, the research instrument was implemented from early May 2023 to mid-June 2023. By the end of June 2023, all data had been collected, cleaned, and analysed.

3.3 Data Analysis

The proposed research model underwent statistical analysis using SPSS V.29 and LISREL V.8.72 software. The analysis consisted of two stages. In the first stage, Confirmatory Factor Analysis (CFA) was conducted to assess the reliability and validity of the measurement model by examining the relationships between items and constructs. This study utilized several Goodness-of-Fit (GFI), including the Normed Fit Index (NFI), Comparative Fit Index (CFI), Adjusted Goodness-of-Fit Index (AGFI), and Root Mean Square Error of Approximation (RMSEA). Structural Equation Modelling literature suggests that a model demonstrates an excellent fit when NFI, CFI, GFI, and AGFI values exceed 0.95. For RMSEA, values below 0.05 indicate an excellent fit, while values below 0.08 are considered acceptable (Hair et al., 2018). The second stage involved evaluating the structural model, which included assessing the model fit, examining the research hypotheses, and exploring potential moderator effects. These analyses provide insights into the relationships and significance of the variables in the research model.

4. Result and Findings

4.1 Demographic of Samples

The study participants in Table 2 included 911 respondents, 558 females (61.3%) and 353 males (38.7%). The education level, most of the group are 806 bachelor's degree students (88.5%), followed by 65 master's degree students (7.1%), 24 Doctoral degree students (2.6%), and 16 Diploma students (1.8%). Participants were affiliated with various academic departments; the highest group of samples from the Science and Technology department had 483 participants (53.0%). The Social Science and Humanities department had 236 participants (25.9%), while the Healthy Sciences department had 168 participants (18.4%). A smaller number of participants were from the Diploma programme, with 24 participants (2.6%). The average sample age is 21.61 (SD=4.63), the minimum age is 18, and the maximum is 59. The respondents are students in various years; there are 244 freshers (26.8%), 289 sophomores (31.7%), 245 juniors (26.9%), 117 seniors (12.8%), 13 5th-year students (1.4%), and 3 6th year students (0.3%). However, participants who are Generative-AI user has 833 participants (91.4%), and 78 participants are non-Generative-AI users (8.6%).

Table 2: Overall demographic characteristics

Topics	Items	Frequency	%
Gender	Female	558	61.30
	Male	353	38.70
Education Level	Bachelor's degree	806	88.50
	Master's degree	65	7.10
	Doctoral Degree	24	2.60
	Diploma	16	1.80
Department	Science and Technology	483	53.00
	Social Science and Humanities	236	25.90
	Healthy Sciences	168	18.40
	Diploma	24	2.60

Topics	Items	Frequency	%
Year of Study	1st	244	26.80
	2nd	289	31.70
	3rd	245	26.90
	4th	117	12.80
	5th	13	1.40
	6th	3	0.30
Using Generative-AI	Yes	833	91.40
	No	78	8.60

Table 3 shows that 445 participants (53%) were not members of relevant professional organisations or associations, while 388 participants (47%) paid for membership registration. Regarding the type of Generative-AI utilised, the most common category was text Generative (f=435, 47.7%), followed by code Generative (f=240, 26.3%), Image Generative (f=108, 11.9%), sound Generative (f=13, 1.4%), and VDO Generative (f=37, 4.1%). Furthermore, participants reported varying frequencies of using Generative-AI in their educational activities. The most frequent use was "More than 1 time per day" (f=247, 27.1%), followed by "1 time a week" (f=234, 25.7%), "1 time a day" (f=212, 23.3%), "1 time bi-weekly" (f=78, 8.6%), and "1 time a month" (f=62, 6.8%). Regarding the time spent using Generative-AI, most participants (f=323, 35.5%) reported spending 2-4 hours on Generative-AI activities. Other time ranges included 1-2 hours (f=232, 25.5%), 4–7 hours (f=137, 15.0%), less than 1 hour (f=88, 9.7%), and more than 7 hours (f=53, 5.8%). Participants identified numerous benefits derived from using Generative-AI in education. The most reported benefit was "Seeking information" (f=601, 50.89%), followed by "Doing Task/Assignment" (f=208, 17.61%), "Being Pals" (f=172, 14.56%), "Exchanging Ideas" (f=116, 9.82%), and "Entertainment" (f=84, 7.11%). Participants indicated various sources of reference when using Generative-AI. Most participants (f=590, 52.26%) referred to Social Media platforms for information. Friends/Acquaintances were another commonly mentioned source, with 297 participants (26.31%) relying on them. TV served as a reference for 116 participants (10.27%), while Online Video Platforms were used by 126 participants (11.16%).

Table 3: Generative-AI users' behaviour

Topics	Items	Frequency	%
Member of Generative-AI services	Not Member	445	53.00
	Member	388	47.00
Type of Generative-AI	Text Generative	435	47.70
	Code Generative	240	26.30
	Image Generative	108	11.90
	Sound Generative	13	1.40
	Video Generative	37	4.10
Frequency of Use	More than 1 times per day	247	27.10
	One time a week	234	25.70
	One time a day	212	23.30
	One time for bi-weekly	78	8.60
	One time a month	62	6.80
Time to use	2-4 hrs.	323	35.50
	1-2 hrs.	232	25.50
	4-7 hrs.	137	15.00
	Less than 1 hr.	88	9.70
	more than 7 hours	53	5.80
Perceived benefits	Seeking information	601	50.89
	Doing Task/ Assignment	208	17.61

Topics	Items	Frequency	%
	Being Pals	172	14.56
	Exchanging Ideas	116	9.82
	Entertainment	84	7.11
Sources of reference	Social media	590	52.26
	Friends /Acquaintance	297	26.31
	TV	116	10.27
	Online Video Platform	126	11.16

4.2 Measurement Model

In the assessment of our measurement model, the constructs demonstrated varying degrees of fit. Attitude Toward Using technology (AT) showed a satisfactory fit with a Chi-Square of 7.49 (df=5), an RMSEA of 0.092, and good fit indices including CFI=0.97, NFI=0.94, and GFI=0.95. The Perceived Ease of Use (PEOU) and Behavioural Intention (BI) constructs also indicated good model fits, with PEOU recording a Chi-Square of 9.23, RMSEA of 0.120, and CFI of 0.97, and BI showing a Chi-Square of 8.08, RMSEA of 0.102, and CFI of 0.98. However, the Actual Use (AU) construct exhibited a less satisfactory fit, with a higher Chi-Square value of 13.38, RMSEA of 0.169, and lower fit indices such as CFI=0.93 and NFI=0.90. The Expected Benefits (EB) construct showcased the best model fit across all constructs with a Chi-Square of 3.90, RMSEA of 0.000, and perfect or near-perfect fit indices including CFI=1.00 and NFI=0.98. Perceived Usefulness (PU) showed a moderate fit, with a higher Chi-Square of 19.50, RMSEA of 0.130, and fit indices like CFI=0.96 and NFI=0.93.

In the Confirmatory Factor Analysis (CFA) results, the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity are applied to assess the multivariate normality and the sufficiency of the sample. The significant result of Bartlett's sphericity test ($p < 0.01$), along with a KMO measure of 0.971, both exceeding the threshold of 0.60 (Tabachnick and Fidell, 2019), validated the suitability of the data for factor analysis. Tabachnick and Fidell (2019) also recommended that the univariate skewness should be under 2, and univariate kurtosis should be less than 4 for normal distribution. Despite the skewness values of all items below 2, only item AU1 exhibited kurtosis values exceeding 4. As per existing literature, the minimal acceptable value for CA was determined to be 0.60, and factor loadings should be above 0.60 (Hair et al., 2018).

Table 4 shows convergent and divergent validity were accessed via CFA on the measurement model, which shows that CA and Composite Reliability (CR) tests were performed to assess reliability. CA was applied to measure the internal consistency amongst items as seen in Table 5, while CR was used to describe the extent to which a train of items can represent potential constructs. The values ranged from 0.84 to 0.87, the CR values ranged from 0.44 to 0.90, the Factor Loading (FL) ranged from 0.46 to 0.9, and finally, the values of the Average Validity Extracted (AVE) of variables ranged from 0.52 to 0.90. Fornell and Larcker (1981) suggested that if AVE is less than 0.50 but CR is higher than 0.60, the convergent validity of the construct is still satisfactory.

Table 4: Descriptive statistics and measurement model

Construct	Items	Questions	M	SD	SK	KU	FL	α	CR	AVE
EB	EB1	Do you think artificial intelligence can help assess complex tasks and suggest real-time personalised recommendations for you?	6.08	0.91	-1.34	3.65	0.52	0.88	0.64	0.90
	EB2	Do you think that know-how artificial intelligence will help create smart agents? (Robot or software) to function as a learning partner or teaching assistant in learning?	5.98	0.95	-1.35	3.71	0.86			
	EB3	Do you think using artificial intelligence will help you plan or perform the assignments from teachers with quality?	5.95	0.86	-0.86	2.31	0.83			
	EB4	Do you think using artificial intelligence will increase your chances of improving your academic performance?	5.99	0.92	-0.89	1.77	0.84			
	EB5	Do you think that using artificial intelligence can help classmates see your existing learning abilities?	5.92	1.06	-1.38	3.16	0.90			

Construct	Items	Questions	M	SD	SK	KU	FL	α	CR	AVE
PU	PU1	Do you think that using know-how artificial intelligence makes learning activities easier?	6.11	0.84	-0.85	1.15	0.81	0.86	0.90	0.58
	PU2	Do you think that using know-how artificial intelligence helps you learn quickly?	6.03	0.86	-0.90	1.56	0.71			
	PU3	Do you think that using artificial intelligence will benefit your learning?	6.04	0.83	-0.68	0.57	0.74			
	PU4	Do you think that using knowledgeable artificial intelligence to help with your assignments?	6.00	0.90	-0.70	0.58	0.78			
	PU5	Do you think using cognitive artificial intelligence enhances your learning efficiency?	6.01	0.91	-1.06	2.56	0.60			
	PU6	Do you think that using artificial intelligence that you can create will make you more knowledgeable?	5.99	0.89	-0.75	1.05	0.78			
PEOU	PEOU1	Do you think you can quickly learn about using know-how artificial intelligence?	6.09	0.91	-1.25	3.25	0.72	0.84	0.84	0.52
	PEOU2	Do you think that using know-how artificial intelligence does not require much effort?	5.93	0.97	-1.17	2.60	0.72			
	PEOU3	Do you think that using know-how artificial intelligence is simple?	5.95	0.97	-1.32	3.25	0.55			
	PEOU4	Do you think that you can use artificial intelligence without asking for help from others?	5.94	1.00	-1.30	3.41	0.75			
	PEOU5	Do you think that using know-how artificial intelligence is easy for you?	5.98	0.94	-0.88	1.10	0.81			
AT	AT1	Do you think that it is a good thing to use artificial intelligence?	6.11	0.84	-0.66	0.25	0.66	0.81	0.50	0.83
	AT2	Do you know how to use Artificial Intelligence to enhance your learning?	6.04	0.81	-0.62	0.64	0.72			
	AT3	Do you think cognitive Artificial Intelligence is valuable to your learning and education?	6.05	0.85	-0.81	1.42	0.88			
	AT4	Do you feel comfortable incorporating artificial intelligence into your learning?	5.96	0.88	-0.64	0.90	0.60			
	AT5	Do you think that know-how artificial intelligence makes work easier and faster?	6.05	0.86	-0.62	0.22	0.62			
BI	BI1	When encountering a problem, do you think you will use artificial intelligence before asking others?	5.86	1.20	-1.64	3.58	0.82	0.86	0.86	0.56
	BI2	Do you think that when you encounter a problem, you will only use artificial intelligence to help solve it?	5.67	1.38	-1.66	2.86	0.90			
	BI3	Do you think that artificial intelligence can solve problems better than humans?	5.74	1.19	-1.40	2.79	0.74			
	BI4	Do you think artificial intelligence knows how to create and solve problems? You can work as you want.	5.79	1.04	-1.01	1.95	0.71			
	BI5	Do you think you need artificial intelligence to be developed more responsive to your lifestyle?	6.00	0.88	-0.96	2.16	0.50			
AU	AU1	Do you think you always use artificial intelligence to create text, images, or videos?	5.86	1.09	-1.66	4.63	0.46	0.78	0.45	0.80
	AU2	Do you think that you are interested in using artificial intelligence in the future?	6.08	0.81	-0.65	0.66	0.85			
	AU3	Do you think you will use artificial intelligence to support learning every time?	5.97	0.95	-1.31	3.47	0.68			

Construct	Items	Questions	M	SD	SK	KU	FL	α	CR	AVE
	AU4	Do you think that you are willing to keep up to date with your knowledge-based artificial intelligence skills?	6.03	0.93	-0.78	0.81	0.64			
	AU5	Do you think that you are happy to use knowledgeable artificial intelligence to support learning?	5.97	0.90	-0.66	0.83	0.66			

Table 5: Inter-construct correlations

	AU	BI	AT	PU	PEOU	EB
AU	1.00					
BI	0.65	1.00				
AT	0.79	0.57	1.00			
PU	0.80	0.65	0.84	1.00		
PEOU	0.68	0.70	0.67	0.74	1.00	
EB	0.74	0.74	0.72	0.82	0.75	1.00

Note. Values on the diagonal represent Pearson's Correlation value.

4.3 Structural Model: Goodness of fit Statistics and Hypothesis Testing

First, the overall model fit was assessed using multiple fit criteria; seven Goodness-of-Fit indices were used, including Chi-Square/Degree of Freedom, Goodness-of-Fit index (GFI), Adjusted Goodness-of-Fit Index (AGFI), Normalised Fit Index (NFI), Non-Normalised Fit Index (NNFI), Comparative Fit Index (CFI), and Root Mean Square Residual (RMSR). The SEM analysis revealed that the goodness of fit statistics of the theoretical framework in Table 5 represented a good fit (Chi-Square=1183.14, df=401, $p < 0.000$, $\chi 2$ /df=2.95, GFI=0.90, AGFI=0.90, NFI=0.98, NNFI=0.99, CFI=0.99, RMSEA=0.48, SRMR=0.045).

An assessment of the direct effects between the research constructs was performed, and the results were as follows: EB (β=1.15, t=7.46 and β=0.92, t=19.72) has a significant positive influence on PU and PEOU in Generative-AI. PU (β=0.99, t=11.11) showed significant effects on Attitude toward Generative-AI. AT (β=1.02, t=14.01) significantly positively affected Behavioural Intention. Then, BI (β=0.78, t=16.36, $p < 0.00$) is significant in determining the AU of adopting Generative-AI. However, PEOU has no significant positive effect PU (β= -0.18, t= -1.25) and AT toward (β=0.01, t=0.08) in Generative-AI. Thus, most hypotheses were supported, as shown in Table 6 and Figure 2.

Table 6: Structural Equation Modelling results of the proposed model

Hypothesis	Standardized Solution	t-value	Results
H1: EB → PU	1.15	7.46***	Supported
H2: EB → PEOU	0.92	18.72***	Supported
H3: PEOU → PU	-0.18	-1.25	Not Supported
H4: PEOU → AT	0.01	0.08	Not Supported
H5: PU → AT	0.99	11.11***	Supported
H6: AT → BI	1.02	14.01***	Supported
H7: BI → AU	0.99	16.10***	Supported

** $p < 0.01$. *** $p < 0.001$.

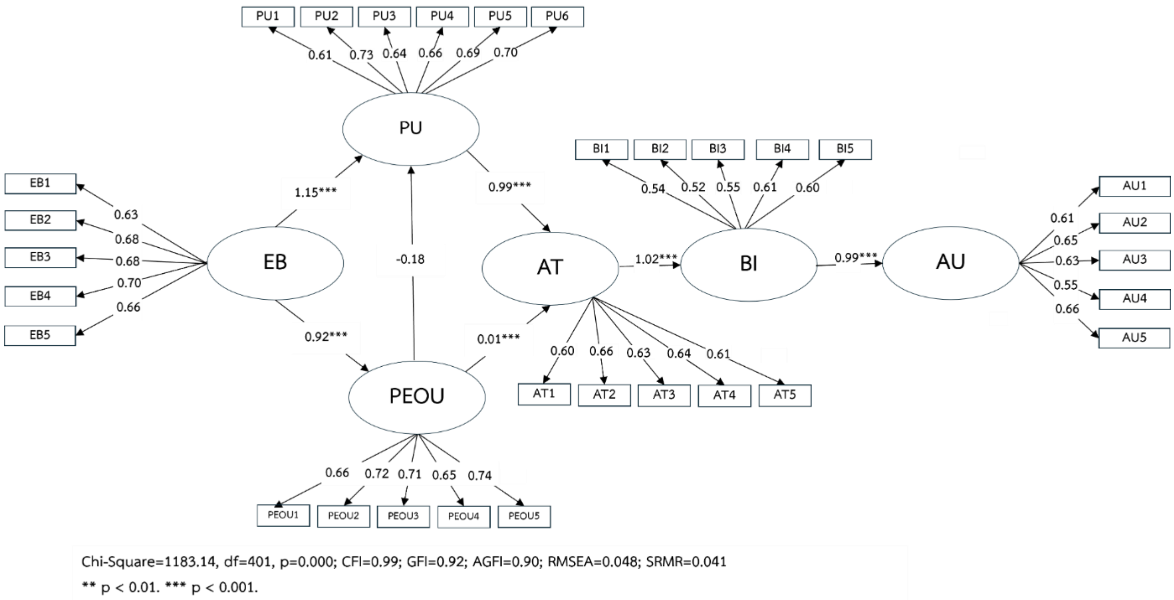


Figure 2: Result of Structural Equation Modelling

5. Discussion and Implications

The expeditious advancement of Generative-AI has the potential to revolutionise higher education, expanding new opportunities for personalised learning, knowledge discovery and transformation. This study contributes to the blooming body of research on AI adoption in education by investigation the factors influencing the acceptance of Generative-AI amongst Thai higher education students through the lens of the TAM.

One of the most captivating findings of this study is the negative relationship between Perceived Ease of Use (PEOU) and Perceived Usefulness (PU), challenging conventional assumptions of TAM (Venkatesh and Davis, 2000). Suggesting that as Generative-AI becomes more intuitive and user-friendly, students may paradoxically perceive it as less valuable for their learning. This could be attributed to the new generation of learners' emerging expectations and digital literacy. More sophisticated adaptive AI tools to keep pace with their learning needs (Ahmed et al., 2021; Keane et al., 2023). With continuous evolution and integration with other emerging technologies like virtual reality and brain-computer interfaces, AI-enable learning experiences that are not only easy to use but also cognitively challenging and emotionally engaging must be provided.

Another condemning implication of this study relates to the role of language proficiency in the even-handed access and adoption of Generative-AI in education. Whilst the current dominance of English in AI systems poses a barrier for non-native speakers (Bulathwela et al., 2021), the expeditious development of multilingual AI models and the increasing availability of municipal languages datasets offers promising solutions for bridging the language divergence. (Padmakumar, Stone and Mooney, 2018). Additionally, opportunities for creating immersive and personalised language acquisition experiences are presented by integrating Generative-AI in language learning as AI-powered language translation and generation become more accurate and contextually aware. Learners from diverse linguistic backgrounds can seamlessly collaborate and learn from each other using Generative-AI as a universal communication tool.

The crucial role of universities in promoting the use and innovative application of advanced content-creation technologies within the educational sector is underlined. Beyond solely consolidating these technologies into their syllabi, higher education institutions are encouraged to cultivate an environment conducive to innovation and experimentation. This involves motivating both students and faculty to collaborate on and explore new uses for these technologies across various academic fields (Chen, Chen and Lin, 2020; Sun et al., 2021). Achieving this goal requires a significant transformation in the educational approach, shifting from a traditional teacher-centric model to one that is centred around the learner. Learners are encouraged to take control of their educational journeys, using advanced tools for creative expression, knowledge searching, solving complex problems, and generating new knowledge.

Additionally, there is a call for the establishment of interdisciplinary research centres and hubs for innovation. These facilities would unite specialists from the fields of computer science, education, psychology, and others to examine the enduring effects of these technologies on learning processes, cognitive development, and societal interactions (Ng et al., 2022; UNESCO., 2023a).

However, the transformative possibility of Generative-AI in education also raises profound ethical and societal questions that cannot be ignored. As AI-generated content becomes increasingly sophisticated and indistinguishable from human-created works resulting in a risk of blurring the boundaries between originality, plagiarism, creativity, and automation (Su and Yang, 2023), educators and policymakers must proactively develop ethical frameworks and guidelines for the responsible use of Generative-AI in education to ensure that learners are well equipped with moral compass along with the critical thinking and knowledge creation skills (Wang et al., 2023). Moreover, the widespread adoption of Generative-AI in education may provoke existing inequalities and create new forms of digital divide, as learners from disadvantaged backgrounds may lack the access, skills, and support needed to leverage these technologies for their learning. These challenges will a concerted effort from all stakeholders, including educators, technology developers, civil society organisations and the governments to ensure an equitable access to Generative-AI in education and that no student is left behind (Ng et al., 2022).

In conclusion, this study provides a subtle understanding of the factors shaping the adoption of Generative-AI in Thai higher education. Embracing the opportunities and challenges of Generative-AI in education is a harness of its potential to create more engaging and personalised learning experiences for all learners while fostering the skills and values needed for responsible citizenship in AI-powered society.

6. Limitations and Future Studies

The present study provides valuable insights into the factors influencing the adoption of Generative-AI amongst learners of higher education in Thailand. Nevertheless, it is imperative to acknowledge the limitations and identify avenues for future research.

A primary limitation resides in the study's focus on a specific population: Thai learners enrolled in Thai universities. To enhance the understanding of Generative-AI adoption in education, future studies should encompass a more diverse sample, incorporating secondary students, teachers, and instructors from varied educational contexts and cultural backgrounds. This broader perspective would facilitate a more nuanced analysis of how different educational settings and user characteristics may influence the acceptance and utilisation of Generative-AI tools.

Another limitation of this study is its cross-sectional design, which collects data at a single point in time. Longitudinal studies would provide valuable insights into how user attitudes and behaviours towards Generative-AI in education evolve over time, given the rapid development in this field. Additionally, qualitative research methods, such as interviews and focus groups, could deepen our understanding of users' experiences, motivations, and challenges when interacting with Generative-AI technologies in educational contexts.

It is also imperative to acknowledge that this study focuses primarily on the Technology Acceptance Model (TAM) as its theoretical framework. While TAM offers a valuable lens for comprehending the key factors influencing technology adoption, future research could benefit from integrating additional theoretical perspectives, such as the Unified Theory of Acceptance and Use of Technology (UTAUT) or the Diffusion of Innovations Theory. These frameworks could assist in capturing a broader spectrum of individual, social, and contextual factors that shape the adoption and utilisation of Generative-AI in education.

Moreover, this research did not thoroughly examine the potential moderating effect of demographic variables, including gender, age, or academic discipline, on the interactions amongst the TAM constructs. Future research could investigate how these elements influence students' perceptions and behaviours regarding Generative-AI, offering a more intricate understanding of adoption patterns. It can also inform targeted interventions suitable for various user populations.

Lastly, in the context of Generative-AI in education, further research is necessary to explore the pedagogical implications and optimal strategies for integrating these tools into teaching and learning activities. Future studies could examine how Generative-AI can be effectively harnessed to facilitate various educational goals, such as personalised learning, collaborative problem-solving, and creative expression. Additionally, research on the ethical and societal aspects of Generative-AI in education, addressing concerns about privacy, bias, and intellectual property, is crucial. This will guide the responsible development and implementation of these technologies.

In conclusion, while this study contributes significantly to understanding Generative-AI adoption in Thai higher education, it also emphasises the necessity for more research to overcome its limitations and explore new directions. By building upon the findings of this study and expanding its scope, future research can contribute to unlocking the full potential of Generative-AI in education. It can also ensure the development and utilisation of these technologies in an equitable, ethical, and pedagogically sound manner.

Author Contributions

Conceptualization, Pawarit Pingmuang, Suchaya Wisnuwong, Noawanit Songkram, and Jintavee Khlaisang; *methodology*, Noawanit Songkram; *software*, Thewawuth Simasathien; *validation*, Thewawuth Simasathien; *formal analysis*, Thewawuth Simasathien; *investigation*, Pawarit Pingmuang and Suchaya Wisnuwong; *resources*, Benz Wiwatsiripong; *data curation*, Benz Wiwatsiripong, Kraisila Kanont, and Kanitta Poonpirome; *writing—original draft preparation*, Kraisila Kanont, Pawarit Pingmuang and Suchaya Wisnuwong; *writing—review and editing*, Suchaya Wisnuwong and Pawarit Pingmuang; *visualization*, Thewawuth Simasathien; and *project administration*, Pawarit Pingmuang

All authors have read and agreed to the published version of the manuscript.

Disclosure statement

All authors declare that they have no conflict of interest.

Acknowledgement

We would like to extend my deepest gratitude to *Professor Noawanit Songkram* and *Professor Jintavee Khlaisang*, Faculty of Education, Chulalongkorn University, whose expert guidance, invaluable support, and insightful feedback have been instrumental throughout the entirety of this research journey.

References

Ahmad, S. F., Rahmat, M. K., Mubarik, M. S., Alam, M. M., and Hyder, S. I., 2021. Artificial Intelligence and Its Role in Education. *Sustainability (Switzerland)*, *[e-journal] 13*(22). https://doi.org/10.3390/su132212902

Ahmed, Y. A., Mohamed, M. M., Ali, A. F., Alasso, M. M., Siyad, A. D., and Ahmad, M. N. 2021. Evaluating Students Perspectives on ICT Readiness in Somali Higher Education towards Teaching - Learning Acceptance. https://doi.org/abs/2108.11455

Al-Adwan, A. S., Li, N., Al-Adwan, A., Abbasi, G. A., Albelbis, N. A., and Habibi, A., 2023. Extending the Technology Acceptance Model (TAM) to Predict University Students' Intentions to Use Metaverse-Based Learning Platforms. *Education and Information Technologies*. [e-journal]. https://doi.org/10.1007/s10639-023-11816-3

Al-Emran, M., Mezhuyev, V., and Kamaludin, A. ,2018. Technology Acceptance Model in M-learning context: A systematic review. *Computers & Education*, [e-journal] 125, 389-412. https://doi.org/10.1016/j.compedu.2018.06.008

Al Darayseh, A., 2023. Acceptance of artificial intelligence in teaching science: Science teachers' perspective. *Computers and Education: Artificial Intelligence*, *[e-journal] 4*. https://doi.org/10.1016/j.caeai.2023.100132

Buddemeyer, A., Walker, E., and Alikhani, M., 2021. Words of Wisdom: Representational Harms in Learning From AI Communication. https://doi.org/10.48550/arXiv.2111.08581

Bulathwela, S., Pérez-Ortiz, M., Holloway, C., and Shawe-Taylor, J., 2021. Could AI Democratise Education? Socio-Technical Imaginaries of an EdTech Revolution. https://doi.org/10.48550/arXiv.2112.02034

Cao, G., Duan, Y., Edwards, J. S., and Dwivedi, Y. K., 2021. Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making. [e-journal] *Technovation*, *106*. https://doi.org/10.1016/j.technovation.2021.102312

Celik, I., Dindar, M., Muukkonen, H., and Järvelä, S., 2022. The Promises and Challenges of Artificial Intelligence for Teachers: a Systematic Review of Research. *TechTrends*, *[e-journal] 66*(4), pp.616-630. https://doi.org/10.1007/s11528-022-00715-y

Chan, C. K. Y., and Hu, W., 2023. Students' Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education. *Computers and Society*. *[e-journal].* https://doi.org/10.48550/arXiv.2305.00290

Chatterjee, S., and Bhattacharjee, K. K., 2020. Adoption of artificial intelligence in higher education: a quantitative analysis using structural equation modelling. *Education and Information Technologies*, *[e-journal] 25*(5), pp.3443-3463. https://doi.org/10.1007/s10639-020-10159-7

Chen, L., Chen, P., and Lin, Z., 2020. Artificial Intelligence in Education: A Review. *IEEE Access*, *[e-journal] 8*, pp.75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chou, C.-M., Shen, T.-C., Shen, T.-C., and Shen, C.-H., 2022. Influencing factors on students' learning effectiveness of AI-based technology application: Mediation variable of the human-computer interaction experience. *Education and Information Technologies*, *[e-journal] 27*(6), pp.8723-8750. https://doi.org/10.1007/s10639-021-10866-9

Chuenphitthayavut, K., Zihuang, T., and Zhu, T., 2020. The prediction of behavioral intention to use online mental health interventions. *Psych J*, *[e-journal] 9*(3), pp.370-382. https://doi.org/10.1002/pchj.333

Collier JE., 2020. *Applied Structural Equation Modeling Using AMOS: Basic to Advanced Techniques*. 1st Ed. Abingdon, Oxon, UK: Routledge. 366 p.

Cooper, G., 2023. Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*, *[e-journal] 32*(3), pp.444-452. https://doi.org/10.1007/s10956-023-10039-y

Cruz-Benito, J., Sánchez-Prieto, J. C., Therón, R., and García-Peñalvo, F. J., 2019. Measuring Students' Acceptance to AI-Driven Assessment in eLearning: Proposing a First TAM-Based Research Model. In P. Zaphiris and A. Ioannou, *Learning and Collaboration Technologies. Designing Learning Experiences* Cham.

Dai, Y., Liu, A., and Lim, C. P., 2023. *Reconceptualizing ChatGPT as a student-driven innovation in higher education*. [pdf] Available at: https://doi.org/10.13140/RG.2.2.33039.05283

Davis, F., Bagozzi, R., and Warshaw, P., 2006. Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. *Journal of Applied Social Psychology*, *[e-journal] 22*, pp.1111-1132. https://doi.org/10.1111/j.1559-1816.1992.tb00945.x

Davis, F. D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *[e-journal] 13*, pp.319–340. https://doi.org/10.2307/249008

Dhingra, M., and Mudgal, R. K., 2019. Applications of Perceived Usefulness and Perceived Ease of Use: A Review. 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART),

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., . . . Williams, M. D., 2021. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *[e-journal] 57*. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Fornell, C., and Larcker, D. F., 1981. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *[e-journal] 18*(1), pp.39-50. https://doi.org/10.2307/3151312

Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Mädche, A., Röglinger, M., Ruiner, C., Schoch, M., Schoop, M., Urbach, N., and Vandirk, S., 2023. *Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education: A Guide for Students and Lecturers*. [pdf] Available at: https://digital.uni-hohenheim.de/fileadmin/einrichtungen/digital/Generative_AI_and_ChatGPT_in_Higher_Education.pdf

Gustafson, T., 2023. *Generative AI – What Could it Mean for Higher Education?* [online] Available at: https://teaching.nmc.edu/generative-ai-what-could-it-mean-for-higher-education/

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E., 2018. *Multivariate data analysis* (8th ed.). Cengage.

Hutson, J., Jeevanjee, T., Graaf, V. V., Lively, J., Weber, J., Weir, G., Arnone, K., Carnes, G., Vosevich, K., Plate, D., Leary, M., and Edele, S., 2022. Artificial Intelligence and the Disruption of Higher Education: Strategies for Integrations across Disciplines. *Creative Education*, *[e-journal] 13*(12), pp.3953–3980. https://doi.org/10.4236/ce.2022.1312253

Jiayang Wu, W. G., Zefeng Chen, Shicheng Wan, Hong Lin. (2023). AI-Generated Content (AIGC): A Survey. https://doi.org/arXiv:2304.06632v1

Kashive, N., Powale, L., and Kashive, K., 2021. Understanding user perception toward artificial intelligence (AI) enabled e-learning. *The International Journal of Information and Learning Technology*, *[e-journal] 38*(1), pp.1-19. https://doi.org/10.1108/IJILT-05-2020-0090

Keane, T., Linden, T., Hernandez-Martinez, P., Molnar, A., and Blicblau, A., 2023. Digital technologies: students' expectations and experiences during their transition from high school to university. *Education and Information Technologies*, *[e-journal] 28*(1), pp.857-877. https://doi.org/10.1007/s10639-022-11184-4

Kelly, S., Kaye, S.-A., and Oviedo-Trespalacios, O., 2023. What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, *[e-journal] 77*,. https://doi.org/10.1016/j.tele.2022.101925

Khlaisang, J., Teo, T., and Huang, F., 2019. Acceptance of a flipped smart application for learning: a study among Thai university students. *Interactive Learning Environments*, *[e-journal],* https://doi.org/10.1080/10494820.2019.1612447

Kim, J., Merrill, K., Xu, K., and Sellnow, D. D., 2020. My Teacher Is a Machine: Understanding Students' Perceptions of AI Teaching Assistants in Online Education. *International Journal of Human–Computer Interaction*, *[e-journal] 36*(20), pp.1902-1911. https://doi.org/10.1080/10447318.2020.1801227

Li, K., 2023. Determinants of College Students' Actual Use of AI-Based Systems: An Extension of the Technology Acceptance Model. *Sustainability (Switzerland)*, *[e-journal] 15*(6). https://www.mdpi.com/2071-1050/15/6/5221

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., and Pechenkina, E., 2023. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators [Article]. *International Journal of Management Education*, *[e-journal] 21*(2). https://doi.org/10.1016/j.ijme.2023.100790

Liu, G. L., Darvin, R., and Ma, C. ,2024. Exploring AI-mediated informal digital learning of English (AI-IDLE): a mixed-method investigation of Chinese EFL learners' AI adoption and experiences. Computer Assisted Language Learning, [e-journal], https://doi.org/10.1080/09588221.2024.2310288

Liu, S., Chen, Y., Huang, H., Xiao, L., and Hei, X., 2018. Towards Smart Educational Recommendations with Reinforcement Learning in Classroom. 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE),

Ma, X., and Huo, Y. ,2023. Are users willing to embrace ChatGPT? Exploring the factors on the acceptance of chatbots from the perspective of AIDUA framework. *Technology in Society*, [e-journal] 75,. https://doi.org/10.1016/j.techsoc.2023.102362

Murugesan, S., and Cherukuri, A. K., 2023. The Rise of Generative Artificial Intelligence and Its Impact on Education: The Promises and Perils. *[e-journal] 56*(05), pp.116-121. https://doi.org/10.1109/mc.2023.3253292

Nazaretsky, T., Cukurova, M., and Alexandron, G., 2022. *An Instrument for Measuring Teachers' Trust in AI-Based Educational Technology* LAK22: 12th International Learning Analytics and Knowledge Conference, Online, USA. Ng, D. T. K., Leung, J. K. L., Su, M. J., Yim, I. H. Y., Qiao, M. S., and Chu, S. K. W. ,2022. *AI Literacy for All*. In AI Literacy in K-16 Classrooms (pp. 21-29). Springer International Publishing. https://doi.org/10.1007/978-3-031-18880-0_3

Ng, D. T. K., Leung, J. K. L., Su, M. J., Yim, I. H. Y., Qiao, M. S., and Chu, S. K. W. ,2022. *AI Education and AI Literacy*. In AI Literacy in K-16 Classrooms (pp. 9-19). Springer International Publishing. https://doi.org/10.1007/978-3-031-18880-0_2

Nguyen, N., 2023. *Opportunities and challenges of AI in higher education*. [online] Available at: https://feedbackfruits.com/blog/opportunities-and-challenges-of-ai-in-higher-education

Nugroho, M. A., Dewanti, P. W., and Novitasari, B. T., 2018. The Impact of Perceived Usefulness and Perceived Ease of Use on Student's Performance in Mandatory E-Learning Use. 2018 International Conference on Applied Information Technology and Innovation (ICAITI),

Padmakumar, A., Stone, P., and Mooney, R. J., 2018. Learning a Policy for Opportunistic Active Learning. [pdf] Available at: https://doi.org/10.48550/arXiv.1808.10009

Pillai, R., Sivathanu, B., Metri, B., and Kaushik, N., 2023. Students' adoption of AI-based teacher-bots (T-bots) for learning in higher education. *Information Technology & People*, *ahead-of-print*(ahead-of-print). , *[e-journal].* https://doi.org/10.1108/ITP-02-2021-0152

Qureshi, R., Shaughnessy, D., Gill, K. A. R., Robinson, K. A., Li, T., and Agai, E., 2023. Are ChatGPT and large language models "the answer" to bringing us closer to systematic review automation? *Systematic Reviews*, *[e-journal] 12*(1), pp.72. https://doi.org/10.1186/s13643-023-02243-z

Rasul, T., Nair, S., Kalendra, D., Robin, M., Santini, F., Ladeira, W., Sun, M., Day, I., Rather, A., and Heathcote, L., 2023. The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning & Teaching*, , *[e-journal] 6*. https://doi.org/10.37074/jalt.2023.6.1.29

Rosenberg, N., 1983. *Inside the Black Box: Technology and Economics*. Cambridge University Press. [pdf] Available at: https://doi.org/10.1017/CBO9780511611940

Rudra, S., 2023. How Can Generative AI Be Used in Higher Ed?. [online] Available at: https://edtechmagazine.com/higher/article/2023/01/how-can-generative-ai-be-used-higher-ed

Sallam, M., Salim, N., Barakat, M., Al-Mahzoum, K., Al-Tammemi, A., Malaeb, D., Hallit, R., and Hallit, S., 2023. *Validation of a Technology Acceptance Model-Based Scale TAME-ChatGPT on Health Students Attitudes and Usage of ChatGPT in Jordan*. [pdf] Available at: https://doi.org/10.2196/preprints.48254

Salinas-Navarro, D. E., Vilalta-Perdomo, E., Michel-Villarreal, R., and Montesinos, L., 2024. Using Generative Artificial Intelligence Tools to Explain and Enhance Experiential Learning for Authentic Assessment. *Education Sciences*, *[e-journal]* 14(1), https://doi.org/10.3390/educsci14010083

Sandu, N., and Gide, E., 2019. Adoption of AI-Chatbots to Enhance Student Learning Experience in Higher Education in India. 2019 18th International Conference on Information Technology Based Higher Education and Training (ITHET),

Sing, C. C., Teo, T., Huang, F., Chiu, T. K. F., and Xing wei, W., 2022. Secondary school students' intentions to learn AI: testing moderation effects of readiness, social good and optimism. *Educational Technology Research and Development*, *[e-journal] 70*(3), pp.765-782. https://doi.org/10.1007/s11423-022-10111-1

Singh, S. V., and Hiran, K. K., 2022. The Impact of AI on Teaching and Learning in Higher Education Technology. *Journal of Higher Education Theory and Practice*, *[e-journal] 22*(13). https://doi.org/10.33423/jhetp.v22i13.5514

Song, N. ,2024. Higher education crisis: Academic misconduct with generative AI. *Journal of Contingencies and Crisis Management*, *[e-journal]* 32(1). https://doi.org/10.1111/1468-5973.12532

Su , J., and Yang , W., 2023. Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education. *ECNU Review of Education*, *[e-journal] pp.*1-12. https://doi.org/10.1177/20965311231168423

Sun, Z., Anbarasan, M., and Praveen Kumar, D., 2021. Design of online intelligent English teaching platform based on artificial intelligence techniques. *Computational Intelligence*, *37*(3), 1166-1180. https://doi.org/https://doi.org/10.1111/coin.12351

Tabachnick, B., and Fidell, L., 2019. *Using Multivariate Statistics*. 7th ed. Pearson.

Toksha, B., Kulkarni, T., and Gupta, P., 2022. Impact of AI on Teaching Pedagogy and its Integration for Enhancing Teaching-Learning. [pdf] Available at: https://doi.org/10.1201/9781003184157-7

UNESCO., 2023a. *The challenges and opportunities of Artificial Intelligence in education*. [online] Available at: https://www.unesco.org/en/articles/challenges-and-opportunities-artificial-intelligence-education

UNESCO., 2023b. *Ministerial Roundtable on Generative AI in Education*. [online] Available at: https://www.unesco.org/en/articles/ministerial-roundtable-generative-ai-education

Venkatesh, V. ,2000,. Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. Information Systems Research, [e-journal] 11(4), 342-365. http://www.jstor.org/stable/23011042

Venkatesh, V., and Davis, F. D., 2000. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, *[e-journal] 46*(2), pp.186-204. https://doi.org/10.1287/mnsc.46.2.186.11926

Wang, X., Anwer, N., Dai, Y., and Liu, A., 2023. *ChatGPT for design, manufacturing, and education*. [pdf] Available at: https://doi.org/10.13140/RG.2.2.35077.22244

World Economic Forum., 2023. The Growth Summit: Jobs and Opportunity for All. [online] Available at: https://www.weforum.org/events/the-growth-summit-jobs-and-opportunity-for-all-2023
Yue, M., Jong, M. S.-Y., and Dai, Y., 2022. Pedagogical Design of K-12 Artificial Intelligence Education: A Systematic Review. *Sustainability (Switzerland)*, *[e-journal] 14*(23). https://www.mdpi.com/2071-1050/14/23/15620
Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F., 2019. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *[e-journal] 16*(1), pp. 39. https://doi.org/10.1186/s41239-019-0171-0
Zhou, X., and Schofield, L., 2024. Using social learning theories to explore the role of generative Artificial Intelligence (AI) in collaborative learning. *Journal of Learning Development in Higher Education, [e-journal]* (30). https://doi.org/10.47408/jldhe.vi30.1031

Quo Vadis, University? A Roadmap for AI and Ethics in Higher Education

Fernando Castelló-Sirvent[1], Vanessa Roger-Monzó[2] and Ricardo Gouveia-Rodrigues[3]

[1]Department of Business Management, Universitat Politècnica de València, Spain

[2]Department of Language Theory and Communication Sciences, Universitat de València, Spain

[3]NECE-UBI Research Centre for Business Sciences, Universidade da Beira Interior, Covilhã, Portugal

fernando.castello@upv.es (corresponding author)
vanessa.roger@uv.es
rjagr@ubi.pt

https://doi.org/10.34190/ejel.22.6.3267

Abstract: In recent years, academic interest in new developments in Artificial Intelligence (AI) and its ethical challenges in higher education has increased. The new emerging technologies that have become popular among the university community in recent times require an exhaustive study to evaluate their impact on academic integrity and plagiarism. The main stakeholders in higher education (SoTL, educational authorities, and policymakers) must understand the new trends and the most relevant studies to have action guides that preserve academic integrity standards in deploying AI in the university. This research analyzes scientific articles published in high-impact journals indexed in the Journal Citation Reports (JCR) (*n*=254) and carries out a bibliometric study using VOSviewer 1.6.18 and WordStat 2023.1. The Normalized Impact per Document (NID) and per Year (NIY) are studied, and four thematic groups and twelve main themes are identified and discussed, allowing the internal research structure of this field of study to be determined. Based on the findings, a roadmap for implementing AI in higher education is proposed, preserving ethical standards and based on three levels (Micro, Meso, Macro). This study offers practical implications for SoTL, academic authorities, and policymakers. Furthermore, the evidence found allows editors of high-impact journals to advise on unclosed gaps and new research trends and new research trends in the area.

Keywords: AI, Ethics, Higher education, University, SoTL

1. Introduction

The generalization of Artificial Intelligence (AI) represents a technological revolution that is transforming all areas of society (Wang, Sun and Chen, 2023). Academic experts highlight the significant changes that will occur in job profiles and work practices. Even though AI is emerging as an exciting technology, there is still considerable uncertainty about how it can be ethically and productively integrated into today's society (Bearman and Ajjawi, 2023). Specifically, it highlights the evolution of Large Language Models (LLM), which have gained prominence since 2010. Based on the generalization of new uses of transformers, these models are redefining the capabilities of artificial intelligence and its practical implications, which marks a clear distinction from previous AI advances that began in the 1950s (Canchila, et al., 2024).

The recent incorporation of Artificial Intelligence (AI) in the university classroom is modifying the teaching-learning process. The debate on its use in university education continues; practical applications are already beginning to influence some aspects of the educational process, though they still need to be fully integrated into current teaching methods. This transformation promises to empower teachers and students but, at the same time, presents new ethical and pedagogical dilemmas (Adams and Pente, 2023). This study focuses on the perspective of university efficiency of three stakeholders: Scholarship of Teaching and Learning (SoTL), academic authorities, and policymakers. SoTL represents a systematic, scholarly inquiry into student learning aimed at improving the practice of teaching in higher education (Felten, 2013).

On the one hand, this framework integrates rigorous academic methods to explore and enhance both the teaching process and its outcomes, contributing to a scholarly approach to educational practice. On the other hand, SoTL, university authorities, and policymakers need to understand the complexity, opportunities, and limitations presented by applying this new technology, as it is the basis for strengthening the discourse on AI between teachers and researchers (Watanabe, 2022). The university must be critical in preparing students to face the implications of living in an AI-mediated world.

However, recent studies confirm the need for more critical and deep reflection on the pedagogical approaches and ethical risks involved in the application of AI in higher education (Bearman, et al., 2022; Zawacki-Richter, et

Cite this article: Castelló-Sirvent, F., Roger-Monzó, V. and Gouveia-Rodrigues, R. 2024. "Quo Vadis, University? A Roadmap for AI and Ethics in Higher Education", *Electronic Journal of e-Learning*, 22(6), pp 34-51, https://doi.org/10.34190/ejel.22.6.3267

al., 2019). The rapid diffusion of this technology and the consequences of its use force us to clarify the evolution of the academic debate in the field of study. Academics need a clear view of challenges and opportunities to help us understand critical issues and their interconnectedness and analyze hot articles and hot journals. In addition, the SoTL, university authorities, and policymakers must be aware of the framework offered by the literature on AI and academic integrity to facilitate opportunities in new teaching-learning processes and address the main challenges of this technology in higher education.

The heterogeneity of perspectives the literature addresses and the absence of an integrated vision make us wonder about the coherence of policies on AI at the university and its adoption in higher education. In the title of the article, we urge higher education institutions ("Quo Vadis, University?"), to contribute to the clarity of ideas, their aggregation from three levels (Micro, Meso, and Macro), and obtaining an overall purpose that facilitates the adoption and deployment of AI in the university.

This article provides a new approach to AI's challenges and opportunities for Higher Education, overcoming a vision focused on specific uses and applications of AI beyond partial and incomplete debates. This study integrates an aggregate vision that allows overcoming different partial visions in the academic discussion. To this purpose, gaps not closed by the literature and some necessary new theoretical developments are discussed. Formulating necessary future lines of action suggests meaningful publication opportunities for scholars and journal editors while establishing the bases for discussion on an effervescent field that can guide policymakers and university authorities in policy design and the SoTLs in their development control and improvement.

This research aims to clarify the ethical uses of artificial intelligence in higher education and propose a roadmap that facilitates the implementation of AI in higher education, guaranteeing the preservation of academic integrity and ethical standards. A relevant contribution of this study is to guide and advise SoTL, university authorities, and policymakers.

This article is structured as follows: First, the theoretical framework is reviewed. Second, the materials and methods of analysis are reported. Third, the results are presented and discussed (trend, clusters, and main topics; academic efficiency; hot articles and hot journals), emphasizing identified publication opportunities. Fourth, the discussion of the findings is specified in a roadmap for stakeholders that is structured into three levels (Micro, Meso, Macro). Finally, some conclusions are expanded, focusing on the study's limitations and future lines of research.

2. Theoretical Framework

New developments in AI have burst into higher education, becoming a powerful agent of change that provides promising opportunities and proactive changes. However, it will also alter the implementation of AI in higher education, altering established conventions, and it is necessary for teachers and researchers to adapt to its benefits and drawbacks. This type of scientific debate is common when new technologies are introduced in education (Qadir, 2022).

ChatGPT is the most widespread text-generative AI in the classroom (Huang, 2023). Developed based on the OpenAI language model, it performs complex tasks and generates human-like responses, such as summaries or answers to multiple-choice exams (Susnjak, 2022). It's not the only one. Thanks to its nature, driven by deep learning algorithms, other AIs can create texts, such as Gemini (from Google) and digital images, such as ChatGPT 4, among others (Lim, et al., 2023).

However, given AI's remarkable capacity to provide answers and generate content, university teachers point out the potential risks of AI, especially the negative impact on students' academic integrity and ethics (Mhlanga, 2023), weakening of their critical thinking (Susnjak, 2022), and adverse effects on evaluation processes (Chatterjee and Dethlefs, 2023; Rudolph, Tan and Chan, 2023; Stokel-Walker, 2022).

Faced with these threats, two scenarios have been proposed: prohibition and control.

Faced with the fear of the consequences of the developments in AI, various governments and educational institutions have prohibited or limited the use of these technologies (Lim, et al., 2023). However, previous attempts to ban emerging technologies in education have failed (Finkle and Masters, 2014; Spies et al., 2010), so it is likely that a similar situation could occur with AI (Farrokhnia, et al., 2023). The alarm has also been raised in academic publications, and their scientific integrity (Cotton, Cotton and Shipway, 2023; Shiri, 2023) has given rise to reactive regulations (Nature, 2023). Specific software like Scribbr, QuillBot, or ZeroGPT has been used to detect AI-generated content in students' tasks. However, it is proven to be a temporary solution (Rudolph, Tan, and Chan, 2023) since the precision of technological content detection software cannot compete with the improvement of successive AI models (Farrokhnia, et al., 2023). OpenAI's plagiarism detection software (AI Classifier) is no longer operational due to its high rate of inaccuracy (OpenAI, 2023).

At the opposite extreme are higher education institutions, especially in developed countries, that see AI as an ally.

Various research advocates designing learning tasks adapted to the use of these technologies (Farrokhnia, et al., 2023), enhancing skills such as creativity or critical thinking (González-Pérez and Ramírez-Montoya, 2022), and developing personalized and immersive learning experiences (Shen, et al., 2023). It is proposed that the evaluation acts be modified to avoid academic fraud derived from AI. In this sense, it is considered appropriate to resort to formative evaluation that involves students (Rushton, 2005; Banihashem, et al., 2022).

In this context, the Scholarship of Teaching and Learning (SoTL) is essential in guiding teachers in imbricating AI in pedagogical methodologies. The SoTL is a conception of teaching practice to improve the quality of teaching and learning processes in the university environment (Cranton, 2011) that is aimed at "instructors, staff, and learners in developing the necessary skills, knowledge, and behaviors to model and implement strategies that promote academic integrity in their teaching, learning, research, assessment and academic practices" (Kenny and Eaton, 2022, p. 578). SoTL is based on six main principles: (1) Inquiry Focused on Student Learning: SoTL identifies teaching challenges observed directly within the classroom setting. These challenges drive the inquiry, reflecting a practical and immediate relevance to educational practice (Bass, 1999); (2) Grounded in Context: SoTL research is deeply embedded in specific educational contexts, including disciplinary and institutional environments. This grounding ensures that the studies are relevant and applicable locally, enhancing their utility and impact (Felten, 2013); (3) Methodologically Sound: SoTL mandates appropriate and rigorous research methodologies. While the choice of methods may vary across disciplines, they must effectively address the research questions posed (Felten, 2013); (4) Conducted in Partnership with Students: Effective SoTL research involves students as partners. This collaboration not only enriches the learning experience but also embodies a democratic, participatory approach to education (Hutching and Huber, 2005); (5) Appropriately Public: The findings from SoTL research should be publicly shared to contribute to the community's knowledge base and allow for peer evaluation. This principle emphasizes the importance of transparency and communication in academia (Shulman, 2004); (6) Critically Reflective: Researchers are expected to engage in critical reflection concerning their methodologies and outcomes, which is essential for the continuous improvement of teaching practices.

SoTL plays a critical role in elevating teaching to a level of serious academic inquiry. By adhering to rigorous principles and openly sharing findings, SoTL practitioners affirm education's scholarly legitimacy and academic value (Bernstein, 2008). Furthermore, SoTL fosters a more reflective and evidence-based approach to teaching, thereby enhancing the educational experience for instructors and students.

Research on AI-related needs, preferences, and supports can help university authorities identify best practices for integrating this new technology into the curriculum (Hamilton, 2023) and offering a personalized and inclusive learning experience. It is the response of numerous universities to the need for students to develop the appropriate skills to successfully face the future work environment (Watanabe, 2022).

3. Materials and Methods

This research followed the principles and protocols of the PRISMA 2020 statement (Figure 1), commonly used in conducting systematic reviews and meta-analyses on multiple topics (Sarkis-Onofre, et al., 2021; Rethlefsen, et al., 2021), with emphasis on AI in different productive sectors (Regona, et al., 2022; Mustapha, et al., 2023) and AI and university students (Zhai, 2023; Crompton and Burke, 2023; González-Calatayud, Prendes-Espinosa and Roig-Vila, 2021). Based on the established objectives, an analysis of the underlying academic debate in the articles included in the literature review was proposed. Additionally, academic efficiency was analyzed to identify and discuss research trends and directions based on hot articles and hot journals.

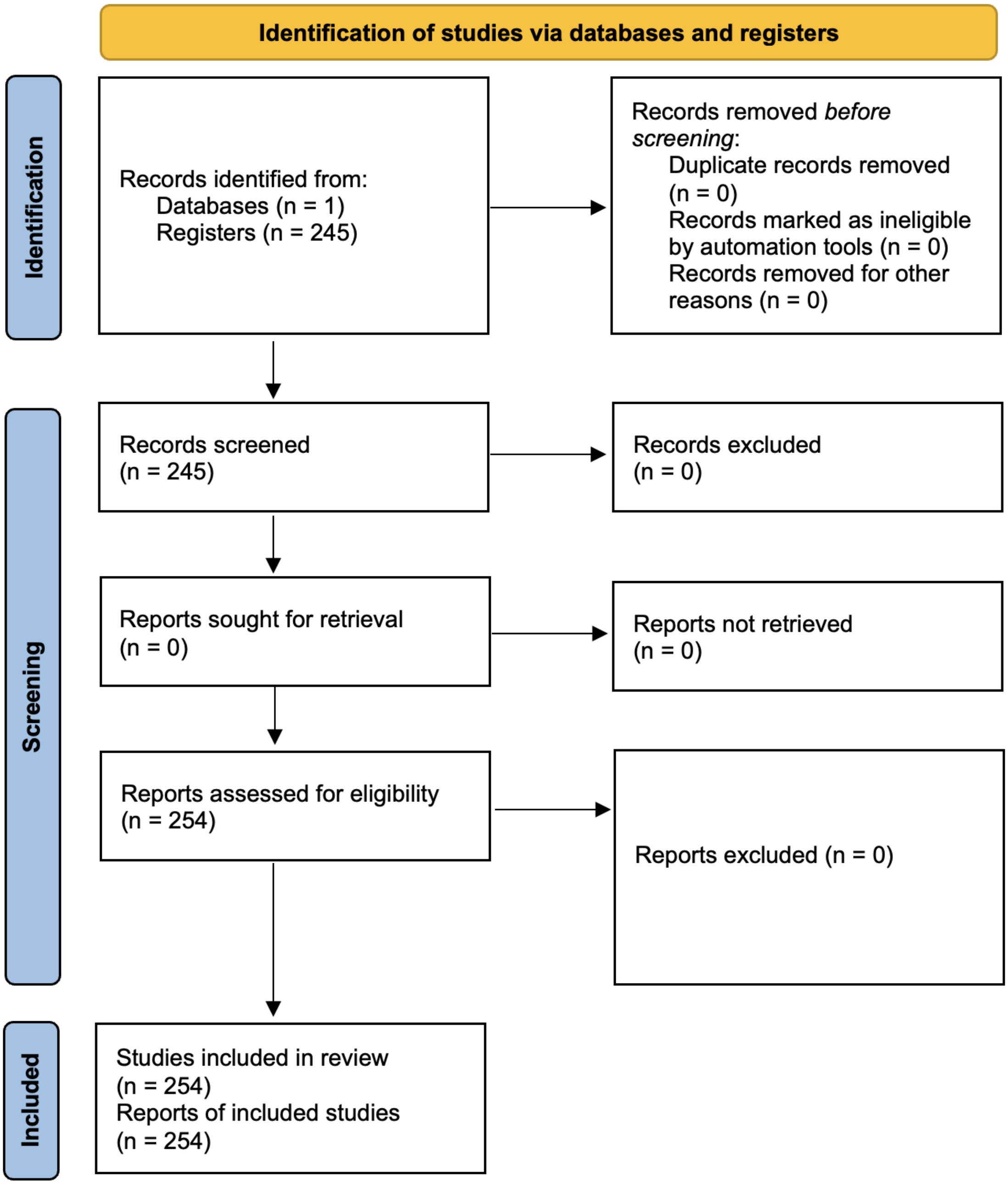


Figure 1: PRISMA statement flow diagram

Web of Science (WoS) was used to create the database for this study. WoS, owned by Clarivate Analytics, is the collection of databases of bibliographical references and citations of scientific publications that collect information from 1900 to the present. In this research, articles published in journals included in the Social Sciences Citation Index (SSCI) and Science Citation Index Expanded (SCIE) were selected, guaranteeing the high impact of the academic literature analyzed since all the articles evaluated in this study are included in the Journal Citation Reports (JCR). The Boolean search string was introduced in WoS as follows: *(AK="artificial intelligence" AND (AK="higher education" OR AK=universit* OR AK="academic integrity" OR AK=plagiarism)) OR (AK=ai AND (AK= "higher education" OR AK=universit * OR AK="academic integrity" OR AK= plagiarism)) OR (TI="artificial intelligence" AND (TI="higher education" OR TI=universit* OR TI="academic integrity" OR TI=plagiarism)) OR (TI=ai AND (TI= "higher education" OR TI=universit* OR TI="academic integrity" OR TI= plagiarism)).*

The bibliometric analysis used VOSviewer v.1.6.18, a software used in scientometric research, and contributes to the visualization of connections and trends (Van Eck and Waltman, 2010). Among many other fields, VOSviewer has been used to study the academic debate on e-learning (Tibaná-Herrera, Fernández-Bajón and Moya-Anegón, 2018) or exploratory research on AI ethics in education (Yu and Yu, 2023). In this study, VOSviewer was used to analyze the co-occurrence of author keywords of the analyzed articles, setting a minimum threshold of three articles to identify the main thematic clusters that underlie the academic debate on AI, university, and academic integrity.

In addition to the keyword analysis with VOSviewer v.1.6.18, the abstracts of the articles are analyzed with WordStat v. 2013.1, a content analysis software that integrates text mining tools and allows topic and trend extraction. Content analysis performed with WordStat complements other statistical and bibliometric software analyses. WordStat has been used in the field of study of insurance sectors (Ellili, et al., 2023), AI in manufacturing (Zeba, et al., 2021) or AI for digital sustainability (Pan and Nishant, 2023), among others.

In this study, WordStat v. 2003.1 was used for the content analysis of the abstracts using an Exploratory Factor Analysis (EFA) using a VARIMAX rotation to observe the variability in the sentences analyzed, according to the eigenvalues and the correlations expressed by the factor loadings of the observed variables (words) with the factors (topics) (Van Haneghan, 2021). The minimum factor loading was determined at the threshold of 0.2, and a segmentation by sentence was followed. In addition, WordStat calculated the weighted average of word correlations according to Normalized Pointwise Mutual Information (NPMI).

The Normalized Impact per Year (NIY) guarantees that the Roadmap resulting from this study allows us to understand the sources that generate academic debate. The NIY identifies articles with a relevant impact in a homogeneous and harmonized way (e.g., above the mean, the median, or in the first quartile of the total distribution of articles: NIY-Q1) (Castelló-Sirvent, 2022). The NIY offers scholars relevant information on the vortexes generating academic debate in a rapidly spreading study area. According to its definition, the NIY considered the total impact of a scientific article from the total number of citations obtained for its calculation. This absolute impact is harmonized when it is related to the number of years elapsed between the article's publication and the bibliometric study (Da Silva, Castelló-Sirvent and Canós-Darós, 2022).

Hot articles are considered to have been published in the last two years from the completion date of a bibliometric study and have an NIY located in the first decile. Under specific circumstances that justify it in a bibliometric particular survey, the requirement criterion can be modified, in more or fewer years, according to the characteristics of the bibliometric study or, particularly, according to the novelty and enthusiasm with which it has been performed and popularized a new topic introduced into the academic debate.

This research considers all the articles published in the last two years with NIY in the first decile as hot articles. Similarly, hot journals have published articles with NIY Average in the first decile of the NIY distribution.

Additionally, the Normalized Impact per Document (NID) can be calculated for all articles published by a journal or by authors from a given institution or country. In this study, the NID is used to apply the academic production published by scholars whose academic affiliation is linked to a particular country. In this way, the NID by country is obtained, which makes it possible to identify and evaluate academic efficiency by country.

4. 4. Results

4.1 Academic Debate

After applying the Boolean search criteria, articles published in high-impact journals indexed in the Journal Citation Reports (JCR) (*n*=254) in the study area were obtained according to the requirements described in the Materials and Methods section. According to the PRISMA statement, the review of the selected articles guaranteed no duplicates or exclusion criteria. Figure 2 shows the intertemporal evolution of articles published since the first available document according to the search criteria established in this research. As of 2020, the publication of articles in the specific area and the barrier of 10 articles per year in JCR journals have been exceeded. Table 1 compares periods (P1: 1961-2019; P2: 2020-2023) for years, articles, citations, and NIY (Average).

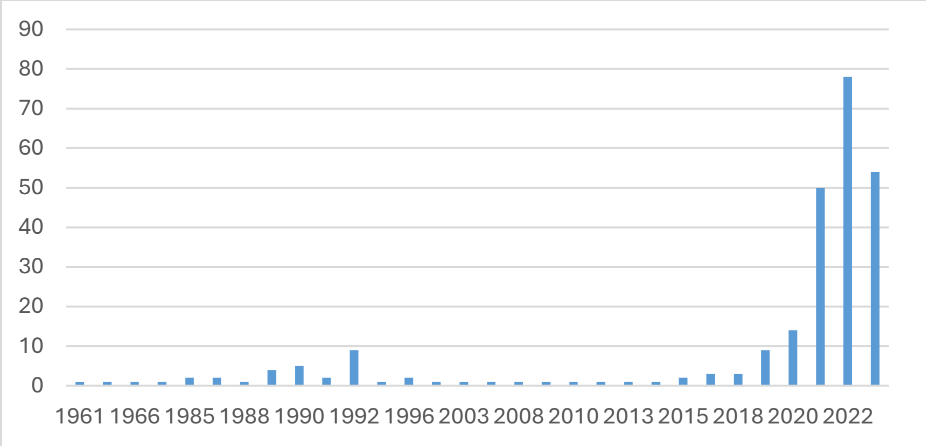


Figure 2: Trend of academic interest in the research area

Table 1 summarizes the articles published according to the threshold set in 2020.

Table 1: Detail analysis

	P1	P2	Total
Period	1961-2019	2020-2023	
Years	59	4	
Articles	58	196	254
Citations	461	779	1,240
NIY (Average)	1.45	1.73	

Note. The articles included in the 2023 data are those available in WoS on 7/31/2023.

NIY = Normalized Impact per Year

The information presented is divided into two periods (P1: 1961-2019; P2: 2020-2023). It shows the total number of articles published, their academic impact measured as the total citation count earned, and the Normalized Impact per Year (NIY) average. The available evidence confirms the trend change observed in Figure 2. The total academic production in 4 years was 3.4 times higher than that published in the previous 59 years. Besides, after 2020, a significant increase in the average NIY is identified.

The results of the Exploratory Factor Analysis (EFA) of the abstracts carried out with WordStat v.2023.1 report 12 topics. Table 2 shows both results.

In the left column, the four clusters identified with VOSviewer v.1.6.18 are presented and connected with the topics reporting on Normalized Pointwise Mutual Information (NPMI), eigenvalue (EI), and frequency (FR), according to the analysis carried out with WordStat v. 2023.1.

Table 2: Main Topics

Cluster	Topic	NPMI	EI	FR
Artificial intelligence on education (red)	Artificial intelligence	0.55	2.55	715
	Higher education institutions	0.51	2.30	452
	Colleges and universities	0.45	2.22	529
	Students academic performance	0.47	1.81	234
	Design methodology approach	0.50	1.50	303
	Effects	0.42	1.36	94
Attitudes and knowledge (green)	Ideological and political courses	0.47	3.52	42
	Results show	0.43	2.01	151
	Decision making	0.39	1.93	54

Cluster	Topic	NPMI	EI	FR
	Research methods	0.48	1.68	231
	College students	0.42	1.65	536
	Thinking skills	0.43	1.63	192
	Colleges and universities	0.45	1.39	421
	Impact of AI for Social work	0.48	1.58	472
	Key factors	0.43	1.54	187
Machine learning and prediction (blue)	Control group	0.34	1.45	84
	Neural network	0.47	1.87	218
	Information services	0.40	1.80	229
	Language and Machine learning	0.43	1.60	222
	Science System	0.40	1.36	185
	Tools and methods	0.43	1.33	144
Technology design and self-efficacy (yellow)	Future development	0.41	1.55	280
	Information	0.42	1.41	160
	Intelligent systems	0.43	1.40	113

Note. Cluster information according to VOSviewer v.1.6.18 and Topic, NPMI, EI, and FR information according to WordStat v.2023.1. NPMI = Normalized Pointwise Mutual Information. EI = Eigenvalue. FR = Frequency

The analysis with VOSviewer v.1.6.18 for the co-occurrences of the authors' keywords reports four clusters according to a threshold of three published articles (Figure 3).

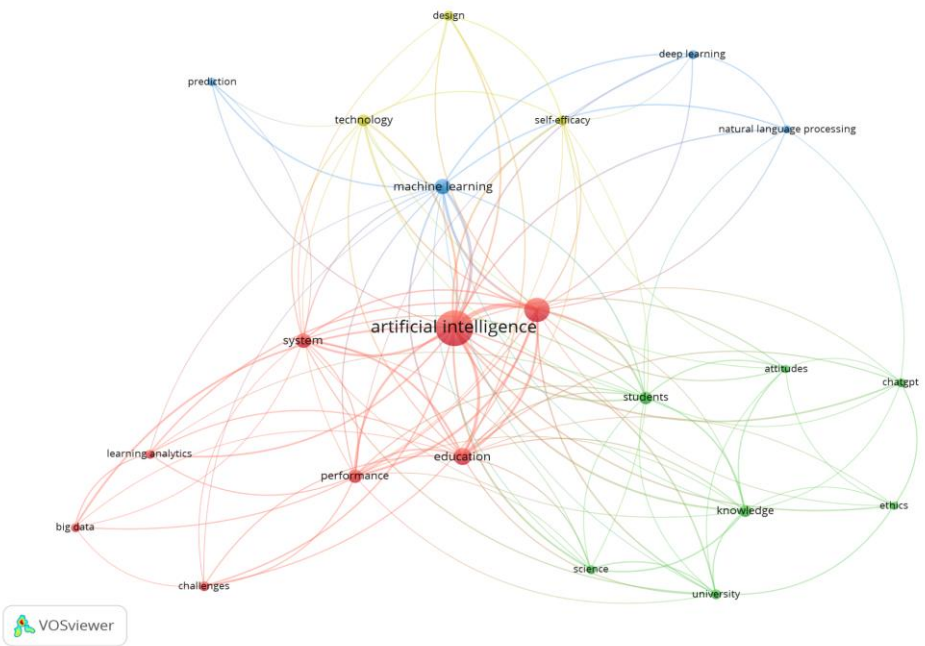


Figure 3: Author Keyword clusters by co-occurrences

Artificial intelligence on education cluster (red) connects relevant research on underlying constraints in the learning environment, such as engaging students with different affective characteristics (Zawacki-Richter, et al., 2019) or secular challenges of higher educational institutions (Chatterjee and Bhattacharjee, 2020). In this cluster, implementations of evaluation of the impact of AI on human agency (Cox, 2021) or big data applications to improve the Student Evaluation of Teaching (SET) massively training Natural Language Processing (NLP) models (Rybinski and Kopciuszewska, 2021) stand out.

Other notable research in the central academic debate on artificial intelligence in education stems from the studies by Deo, et al. (2020) on learning performance prediction in Science, Technology, Engineering, and Mathematics (STEM) teaching, AI analysis on creativity (Wang, Sun and Chen, 2023) or business performance (Khalid, 2020).

Attitudes and knowledge cluster (green) connects studies that use technology-based interventions to change academic integrity (AI) knowledge and attitudes of students (Cronan, et al., 2017) with research that deepens the use of Generative AI (e.g., Chat GPT), as an appropriate tool for developing critical thinking skills and preserving academic integrity (Rusandi, et al., 2023). This cluster also integrates research aimed at delving into the challenges related to making students literate about AI and enhancing their ethical awareness (Kong, Cheung and Zhang, 2023), as well as, in this sense, improving the understanding of some contingent factors and paradoxes that drive and feed the ethical conflict of the use of AI by students (Lim, et al., 2023).

Machine learning and prediction cluster (blue) concentrates research on prediction methods and applications in the context of higher education and university research. Research on student academic performance in online engineering degrees (Jiao, et al., 2022), intelligent libraries (Cox, et al., 2019), or from an approach linked to knowledge transfer in transnational innovation ecosystems according to transnational industries and universities cooperation procedures (TIC, TUC) (Cai, Ramis and Martínez, 2019).

The technology design and self-efficacy cluster (yellow) emphasizes published articles focused on the study of technological and design factors that are facilitators (or inhibitors) of the success in the adoption of AI in the context of higher education, both from the approach from the teachers (Wang, Liu and Tu, 2021) and from the perspective of the students (Almaiah, et al., 2022).

4.2 Academic Efficiency

Table 3 presents the count of countries, citations, and academic efficiency expressed according to the Normalized Impact per Document (NID), assuming a minimum threshold of three articles published in the area.

Table 3: Academic efficiency by countries

Country	Documents	Citations	NID
Germany	4	347	86.8
Serbia	4	61	15.3
Romania	5	66	13.2
Finland	3	36	12.0
Portugal	3	32	10.7
England	11	96	8.7
Canada	4	33	8.3
India	11	88	8.0
Malaysia	6	42	7.0
Indonesia	3	20	6.7
Poland	4	23	5.8
Netherlands	3	15	5.0
Pakistan	3	15	5.0
Usa	22	104	4.7
Jordan	3	14	4.7
Nigeria	3	14	4.7
Mexico	4	17	4.3
Sweden	5	21	4.2
Australia	9	37	4.1

Country	Documents	Citations	NID
Russia	4	15	3.8
Italy	6	21	3.5
Taiwan	12	40	3.3
Saudi Arabia	16	52	3.3
Spain	9	29	3.2
Peoples R. China	92	253	2.8

Note. NID = Normalized Impact per Document

The results reveal high academic efficiency for countries with NID-Q1 scores, such as Germany, Serbia, Romania, Finland, Portugal, England, and Canada. Specifically, Germany and Serbia place their academic efficiency in the first decile (NID > 13.4) of the analyzed distribution. Taiwan, Audi Arabia, Spain, and China have the lowest academic efficiency.

4.3 Hot Articles and hot Journals

According to the results, hot articles are the studies published in 2022 and 2023 with a NIY in the first decile of the distribution studied (NIY > 4.0). Research on the transformation of university education stands out (Okunlava, Syed Abdullah and Alias, 2022; Salas-Pilco and Yang, 2022), AI, plagiarism and honesty in Higher Education (King, 2023; Kleebayoon and Wiwanitkit, 2023), support for students with depression to improve their mental health (Liu, et al., 2022), and medical research (Dahmen, et al., 2023) (For more information, see Appendix; Table 4).

The analysis of hot journals in the area of knowledge evaluated by this study shows the Journals located in the first decile (NIY > 3.2): Cellular and Molecular Bioengineering, Learning Media and Technology, Internet Interventions-The Application of Information Technology in Mental and Behavioural Health, International Journal of Educational Technology in Higher Education, Heliyon, Journal of Medical Internet Research, Knee Surgery Sports Traumatology Arthroscopy, Distance Education, International Journal of Management Education, Library Hi Tech, Education and Information Technologies, Innovations in Education and Teaching International, and Computer Applications in Engineering Education (For more information, see Appendix; Table 5).

5. Discussion

The empirical evidence suggests that scholars have focused their research efforts on AI, universities, and academic integrity on specific applications and use cases, but a complete global vision that addresses the significant challenges that underlie the most immediate future is not offered.

The findings of this study also show that as of 2020, scholarly interest in the impact of the ethical use of AI in higher education has increased drastically, given the need to control the moral integrity of this technology after its widespread use among students. The academic impact measured by the total number of citations and by the NIY average of the articles has been more significant since 2020, which reinforces the existence of an academic debate that is spreading transversally throughout different areas of knowledge beyond the AI and the pedagogical use of this technology (Zawacki-Richter, et al., 2019).

This research offers a Roadmap of AI and academic integrity for SoTL and policymakers, drawing an inspiring map of AI along four main lines of development in the university context:

- education (e.g., effects, academic performance, or design methodology approaches) (Chen, Chen and Lin, 2020)
- attitudes and knowledge (e.g., thinking skills, decision-making, or impacts on social work) (Cox, 2021).
- machine learning and prediction (e.g., language, tools, and methods) (Kuleto, et al., 2021)
- technology design and self-efficacy (e.g., intelligent systems) (Chang, et al., 2022)

Next, the discussion of the implications of this study for SoTL is articulated according to the Micro, Meso, and Macro levels and based on previous taxonomies (Kenny and Eaton, 2022; Poole and Simmons, 2013; Williams, et al., 2013; Poole, 2009).

The Micro-level includes teachers, instructors, and stakeholders who need to identify and address the main lines of self-training and future development trends of the university space in the presence of AI.

The Meso-level includes academic authorities and university managers (e.g., Rectors, Deans, Managers, and other staff members). This type of stakeholder adopts an institutional role and focuses on their institution's organizational and strategic design aspects. They need to articulate coherent actions aligned with the supra-legislative provisions established by policymakers, as determined by their possible frameworks for action. University managers focus their interest on establishing institutional visions at the crossroads of implementation of AI in their universities, as well as the appropriate procedures for adjustment with agents located at the Micro-level.

Politicians, policymakers, and public policy design and evaluation analysts configure the Macro-level. These actors are interested in answering the main questions that arise on the academic horizon and are responsible for designing policies and establishing the regulatory framework that facilitates (or hinders) the adoption of structural changes in AI in universities.

The analysis of the academic literature published in the field of study suggests a broad spectrum of topics yet to be addressed sufficiently and various gaps that must be closed in the coming years. The roadmap articulating the different research lines on this topic is drawn below. The impact of AI and the ethical behavior of students' use of it is a central issue not only for the academic community but also for society. Figure 4 presents a diagram to summarize the roadmap for AI and academic integrity in higher education.

Micro-level
Stakeholders
Teachers Instructors
Roadmap
• Ethical AI in academic quality control • Faculty commitment to academic integrity • Comparing chatbot and human feedback • AI impact on university skill development • Enhancing metacognition and critical thinking • Developing leadership and teamwork skills • Assessing AI exclusion in skill development • Advancing pedagogy for cognitive and soft skills

Meso-level
Stakeholders
Academic authorities University managers
Roadmap
• AI's financial and organizational impact • AI's influence on analytical reasoning • AI for equitable assessment methods • AI training enhances labor market readiness • AI's effect on teacher-student relationships • Evaluating AI skills in academic programs • AI's impact on academic autonomy and freedom

Macro-level
Stakeholders
Politicians Policymakers
Roadmap
• Protecting academic data and IP • Legislative impacts on graduate employability • AI's effect on social mobility • Universities maintain autonomy with tech partnerships • Ensuring ethical AI in democracies

Figure 4: Proposed Roadmap

5.1 Micro-Level

The roadmap proposed in Figure 4 suggests concrete actions for stakeholders who occupy decisive roles at the Micro-level. The proposed Roadmap suggests teachers' attention on: (a) quality control of academic results after the integration of ethical standards in the use of AI by students or kept outside of AI and preserving robot-proof university environments (Aoun, 2017); (b) faculty interest in preserving academic integrity in AI environments; (c) effectiveness of feedback automatically generated by chatbots compared to feedback offered by human instructors; (d) longitudinal and purchased analysis of the impact of the implementation of AI in the university according to the development of different skills (e.g., language or communication), higher cognitive abilities (e.g., metacognition or critical thinking) and soft skills (e.g., leadership or teamwork) and (e) evaluation of the effect of academic responses that keep AI out of the development of skills, higher cognitive skills, and soft skills, calling for new pedagogical approaches (Bearman and Ajjawi, 2023).

5.2 Meso-Level

The proposed roadmap suggests concrete actions for stakeholders who occupy decisive roles at the Meso-level (Figure 4). From the perspective of university authorities and academic managers of faculties and business schools, the tracks that will be consolidated in the coming years in connection with AI, academic integrity and case studies, policies and programs are relevant in accordance with an agenda of emerging research on: (a) financial (cost-benefit) and organizational (interaction-efficiency) impact of AI implementation in university environments; (b) long-term effects on students' analytical reasoning (and as a possible negative externality on industries and society); (c) how AI can provide more fair and equitable evaluation methods, specifically, with minority groups at risk of exclusion, helping to overcome woke ideologies and consequences of anxiety; d) how AI training can better prepare students for a rapidly changing labor market, improving their resilience, especially in the foreseeable context of increasing technological unemployment; (e) how AI could limit the development of interpersonal relationships between teachers and students, negatively impacting engagement and interaction; (f) evaluation of academic programs designed to incorporate skills for the appropriate use of AI, both in students and teachers; (g) impact of AI on academic autonomy and intellectual freedom in university environments from an approach based on affective polarization (Welker, et al., 2023) and cross-cultural moral foundations (Atari, et al., 2023).

5.3 Macro-Level

According to the proposed Roadmap (Figure 4), policymakers must evaluate trends at a Macro-level that allow examining the effects of new legislative frameworks for the development of AI in the university, with emphasis on: (a) privacy, security of personal data and intellectual property of students, teachers and researchers; (b) effects that legislative changes and new public policies could have on the employability of future graduates and their impact on welfare systems; (c) how the development of AI in universities affects social mobility and equality of opportunity, particularly among certain vulnerable groups and in environments of pre-existing poverty; (d) how higher education institutions can preserve their autonomy, freedom of thought and independence by establishing collaboration agreements with large technology companies for the development of AI in universities; (e) how liberal democracies can establish effective international alliances that ensure the ethical and safe development of AI.

6. Conclusions

Scholars' interest in AI, ethics, and university has recently increased. The academic production published in high-impact journals has tripled since 2020 compared to the previous 59 years. Although the evidence suggests that the internal research structure is configured from four clusters (Artificial intelligence on education, Attitudes and knowledge; Machine learning and prediction, Technology design and self-efficacy), the content analysis shows the granularity of the Current academic debate according to twelve main topics. Current research focuses on specific applications of AI in higher education environments, studying specific challenges and opportunities. The research agenda does not include debates that can be extended to a global vision according to a SoTL-based approach. This fragmentation suggests a need for more specificity in the general direction of university policy regarding AI. Throughout the article, we ask ourselves, "Quo Vadis, University?" and the evidence found allows us to configure a roadmap that connects the dots and lets us glimpse what university policy should be like in the emergence of this contingent factor.

This study extends the discussion of the findings and formulates and develops seventeen promising lines of investigation. One of the main contributions of this research is to offer a helpful guide to scholars who wish to direct their research towards emerging topics in AI, ethics, and universities. The study results and discussion provide research opportunities and allow academics to be advised in future research projects. Furthermore, the discussion section proposes a roadmap configured according to three levels (Micro, Meso, and Macro) to integrate AI in higher education from an ethical perspective that guarantees academic integrity.

This article delineates a strategic framework for integrating artificial intelligence (AI) into higher education, focusing on operational levels: Micro, Meso, and Macro. At the Micro-level, educators must enhance academic quality and integrity within AI-enhanced learning environments. This includes evaluating the impact of AI on student performance across various competencies—ranging from technical skills to soft skills like teamwork and leadership. Furthermore, the effectiveness and potential biases of AI-driven feedback versus traditional human feedback must be critically assessed.

Moving to the Meso-level, the roadmap highlights the need for university leaders to consider AI adoption's financial and organizational implications. This includes analyzing AI's influence on student analytical skills and its broader socio-economic repercussions. Universities must develop inclusive evaluation methods that address diversity and fairness, preparing students for dynamic employment landscapes while maintaining interpersonal engagement within the educational process.

At the Macro-level, policymakers are urged to scrutinize AI's legislative and ethical dimensions in academia. This encompasses safeguarding privacy and intellectual property and ensuring equitable social outcomes through education. The roadmap suggests that higher education institutions foster autonomy and collaboration with tech giants, promoting an ethically sound and globally coherent AI integration strategy.

These recommendations aim to foster a balanced, equitable, and forward-thinking adoption of AI in the academic sphere, aligning technological advancements with educational integrity and societal well-being.

This study has several limitations. Firstly, to guarantee the highest standards of academic impact delimited by the research objectives, the search for articles was carried out in journals indexed in JCR. Future articles should expand the databases used, considering repositories that have a faster publication speed even when they sacrifice peer review processes (e.g., Arxiv) to amplify the knowledge base and include updated references in a field of rapidly evolving study.

This research has limited the scope of its recommendations to three stakeholders (professors and instructors, university managers, and policymakers). Future research should complement this approach with other approaches (e.g., the 4M Framework) and the new analytical perspective offered by the stakeholders considered in the Quintuple Innovation Helix Framework. Additionally, a promising avenue of research for scholars arises from the differential analysis of the implementation of AI in the universities of liberal democracies compared to the adoption of AI in other countries far from this secular tradition.

Acknowledgements

This article results from Fernando Castelló-Sirvent and Vanessa Roger-Monzó's research stay as visiting researchers at the NECE-UBI Research Center for Business Sciences of the Universidade da Beira Interior in Covilhã (Portugal).

References

Adams, C. and Pente, P., 2023. Interviewing Artificial Intelligence in Education (AIED): A Posthuman Approach for Learning about AIED's Affordances, Constraints and Ethical Challenges. In: E. Langran, P. Christensen and J. Sanson, ed. 2023. *Proceedings of Society for Information Technology & Teacher Education International Conference*. New Orleans, LA, United States: Association for the Advancement of Computing in Education (AACE), pp.1904-1909. https://doi.org/10.5555/222075

Almaiah, M.A., Alfaisal, R., Salloum, S.A., Hajjej, F., Thabit, S., El-Qirem, F.A. and Al-Maroof, R.S., 2022. Examining the impact of artificial intelligence and social and computer anxiety in e-learning settings: Students' perceptions at the university level. *Electronics*, 11(22), pp.3662. https://doi.org/10.3390/electronics11223662

Aoun, J.E., 2017. *Robot-proof: Higher education in the age of artificial intelligence*. London: MIT Press.

Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S.T. and Dehghani, M., 2023. Morality beyond the WEIRD: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology*, 125(5), pp.1157–1188.https://doi.org/10.1037/pspp0000470

Banihashem, S.K., Noroozi, O., van Ginkel, S., Macfadyen, L.P. and Biemans, H.J.A., 2022. A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 37, 100489. https://doi.org/10.1016/j.edurev.2022.100489

Bearman, M. and Ajjawi, R., 2023. Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, 54, pp.1160-1173. https://doi.org/10.1111/bjet.13337

Bearman, M., Ryan, J. and Ajjawi, R., 2022. Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, 86, pp.369-385. https://doi.org/10.1007/s10734-022-00937-2

Castelló-Sirvent, F., 2022. A fuzzy-set qualitative comparative analysis of publications on the fuzzy sets theory. *Mathematics*, 10(8), 1322. https://doi.org/10.3390/math10081322

Cai, Y., Ramis Ferrer, B. and Martinez Lastra, J.L., 2019. Building university-industry co-innovation networks in transnational innovation ecosystems: Towards a transdisciplinary approach of integrating social sciences and artificial intelligence. *Sustainability*, 11(17), 4633. https://doi.org/10.3390/su11174633

Canchila, S., Meneses-Eraso, C., Casanoves-Boix, J., Cortés-Pellicer, P. and Castelló-Sirvent, F., 2024. Natural language processing: An overview of models, transformers and applied practices. *Computer Science and Information Systems.* https://doi.org/10.2298/CSIS230217031C

Chang, Q., Pan, X., Manikandan, N. and Ramesh, S., 2022. Artificial Intelligence Technologies for Teaching and Learning in Higher Education. *International Journal of Reliability, Quality and Safety Engineering*, 29(5). https://doi.org/10.1142/s021853932240006x

Chatterjee, J. and Dethlefs, N., 2023. This new conversational AI model can be your friend, philosopher, and guide … and even your worst enemy. *Pattern*, 4(1), 100676. https://doi.org/10.1016/j.patter.2022.100676

Chatterjee, S. and Bhattacharjee, K.K., 2020. Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling. *Education and Information Technologies*, 25, pp.3443-3463. https://doi.org/10.1007/s10639-020-10159-7

Chen, L., Chen, P. and Lin, Z., 2020. Artificial Intelligence in Education: A Review. *IEEE Access*, 8, pp.75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510

Cotton, D.R., Cotton, P.A. and Shipway, J.R., 2023. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, pp.1-12. https://doi.org/10.1080/14703297.2023.2190148

Cox, A.M., 2021. Exploring the impact of Artificial Intelligence and robots on higher education through literature-based design fictions. *International Journal of Educational Technology in Higher Education*, 18(1), p.3. https://doi.org/10.1186/s41239-020-00237-8

Cox, A.M., Pinfield, S. and Rutter, S., 2019. The intelligent library: Thought leaders' views on the likely impact of artificial intelligence on academic libraries. *Library Hi Tech*, 37(3), pp.418-435. https://doi.org/10.1108/LHT-08-2018-0105

Cranton, P., 2011. A transformative perspective on the scholarship of teaching and learning. *Higher Education Research & Development*, 30(1), pp.75-86. https://doi.org/10.1080/07294360.2011.536974

Crompton, H. and Burke, D., 2023. Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education*, 20(1), pp.1-22. https://doi.org/10.1186/s41239-023-00392-8

Cronan, T.P., McHaney, R., Douglas, D.E. and Mullins, J.K., 2017. Changing the academic integrity climate on campus using a technology-based intervention. *Ethics & Behavior*, 27(2), pp.89-105. https://doi.org/10.1080/10508422.2016.1161514

Dahmen, J., Kayaalp, M.E., Ollivier, M., Pareek, A., Hirschmann, M.T., Karlsson, J. and Winkler, P.W., 2023. Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. *Knee Surgery, Sports Traumatology, Arthroscopy*, 31(4), pp.1187-1189. https://doi.org/10.1007/s00167-023-07355-6

Da Silva, A.B., Castelló-Sirvent, F. and Canós-Darós, L., 2022. Sensible Leaders and Hybrid Working: Challenges for Talent Management. *Sustainability*, 14(24), 16883. https://doi.org/10.3390/su142416883

Deo, R.C., Yaseen, Z.M., Al-Ansari, N., Nguyen-Huy, T., Langlands, T.A.M. and Galligan, L., 2020. Modern artificial intelligence model development for undergraduate student performance prediction: An investigation on engineering mathematics courses. *IEEE Access*, 8, pp.136697-136724. https://doi.org/10.1109/ACCESS.2020.3010938

Ellili, N., Nobanee, H., Alodat, A.Y., Dilshad, M.N. and Nuzhat, S., 2023. Mapping marine insurance: a bibliometric review: a taxonomical study using bibliometric visualization and systematic analysis. *Journal of Financial Services Marketing*. https://doi.org/10.1057/s41264-023-00232-w

Farrokhnia, M., Banihashem, S.K., Noroozi, O. and Wals, A., 2023. A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), pp.460–474. https://doi.org/10.1080/14703297.2023.2195846

Finkle, T.A. and Masters, E., 2014. Do MOOCs pose a threat to higher education? *Research in Higher Education Journal*, 26, pp.1-10. [Online] Available at: https://files.eric.ed.gov/fulltext/EJ1055324.pdf. [Accessed 23 June 2023].

Garcés-Conejos Blitvich, P., 2022. Moral emotions, good moral panics, social regulation, and online public shaming. *Language & Communication*, 84, pp.61-75. https://doi.org/10.1016/j.langcom.2022.02.002

González-Calatayud, V., Prendes-Espinosa, P. and Roig-Vila, R., 2021. Artificial intelligence for student assessment: A systematic review. *Applied Sciences*, 11(12), 5467. https://doi.org/10.3390/app11125467

González-Pérez, L.I. and Ramírez-Montoya, M.S., 2022. Components of education 4.0 in 21st Century skills frameworks: Systematic review. *Sustainability*, 14(3), 1493. https://doi.org/10.3390/su14031493

Hamilton, M., 2023. Using SoTL Research to Support Educators to Understand How Artificial Intelligence Tools Can Be Effectively Used in the Classroom. In: SSHRC 2023 Public Lightning Talks, *Understanding the Impact of Artificial Intelligence on Higher Education*. Conference Presentation. University of Calgary. 7-9 June 2023.

Huang, K., 2023. Alarmed by AI chatbots, universities start revamping how they teach. *New York Times*. [online] (Last updated 9.47 AM on 1st March 2023. Available at: https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html. [Accessed 23 June 2023].

Jiao, P., Ouyang, F., Zhang, Q. and Alavi, A.H., 2022. Artificial intelligence-enabled prediction model of student academic performance in online engineering education. *Artificial Intelligence Review*, 55(8), pp.6321-6344. https://doi.org/10.1007/s10462-022-10155-y

Kenny, N. and Eaton, S.E., 2022. Academic Integrity Through a SoTL Lens and 4M Framework: An Institutional Self-Study. In: S.E. Eaton, J. Christensen Hughes, ed. 2022. *Academic Integrity in Canada. Ethics and Integrity in Educational Contexts*, vol 1. Cham: Springer International Publishing, pp.573-599. https://doi.org/10.1007/978-3-030-83255-1_30

Khalid, N., 2020. Artificial intelligence learning and entrepreneurial performance among university students: evidence from Malaysian higher educational institutions. *Journal of Intelligent & Fuzzy Systems*, 39(4), pp.5417-5435. https://doi.org/10.3233/JIFS-189026

King, M.R. and ChatGPT, 2023. A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cellular and Molecular Bioengineering*, 16(1), pp.1-2. https://doi.org/10.1007/s12195-022-00754-8

Kleebayoon, A. and Wiwanitkit, V., 2023. Artificial intelligence, chatbots, plagiarism and basic honesty: comment. *Cellular and Molecular Bioengineering*, 16(2), pp.173-174. https://doi.org/10.1007/s12195-023-00759-x

Kong, S.C., Cheung, W.M.Y. and Zhang, G., 2023. Evaluating an artificial intelligence literacy programme for developing university students' conceptual understanding, literacy, empowerment and ethical awareness. *Educational Technology & Society*, 26(1), pp.16-30. Available at: https://www.jstor.org/stable/48707964 [Accessed 26 June 2023].

Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O., Păun, D. and Mihoreanu, L., 2021. Exploring Opportunities and Challenges of Artificial Intelligence and Machine Learning in Higher Education Institutions. *Sustainability*. https://doi.org/10.3390/su131810424

Land, R. and Gordon, G., 2013. *Enhancing quality in higher education: International perspectives*. London: Routledge.

Lim, W.M., Gunasekara, A., Pallant, J.L., Pallant, J.I. and Pechenkina, E., 2023. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Liu, H., Peng, H., Song, X., Xu, C. and Zhang, M., 2022. Using AI chatbots to provide self-help depression interventions for university students: a randomized trial of effectiveness. *Internet Interventions*, 27, 100495. https://doi.org/10.1016/j.invent.2022.100495

Mhlanga, D., 2023. Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4354422

Mustapha, I., Ali, M., Khan, N. and Sikandar, H., 2023. The Impact of Industry 4.0 on Innovative Organisations, A Thematic Review Using the PRISMA Statement 2020. *International Journal of Interactive Mobile Technologies,* 17(9). https://doi.org/10.3991/ijim.v17i09.39465

Nature, 2023. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, 613(7945), 612. https://doi.org/10.1038/d41586-023-00191-1

Okunlaya, R.O., Syed Abdullah, N. and Alias, R.A., 2022. Artificial intelligence (AI) library services innovative conceptual framework for the digital transformation of university education. *Library Hi Tech*, 40(6), pp.1869-1892. https://doi.org/10.1108/LHT-07-2021-0242

Pan, S.L. and Nishant, R., 2023. Artificial intelligence for digital sustainability: An insight into domain-specific research and future directions. *International Journal of Information Management*, 72, 102668. https://doi.org/10.1016/j.ijinfomgt.2023.102668

Poole, G. and Simmons, N., 2013. The Contributions of the Scholarship of Teaching and Learning to Quality Enhancement in Canada. In: G. Gordon, and R. Land, ed. 2013. *Quality Enhancement in Higher Education: International Perspectives*. London: Routledge, pp.118-128. https://doi.org/10.5688/ajpe7408152

Poole, G., Taylor, L. and Thompson, J., 2007. Using the Scholarship of Teaching and Learning at Disciplinary, National and Institutional Levels to Strategically Improve the Quality of Post-secondary Education. *International Journal for the Scholarship of Teaching and Learning*, 1(2), pp.1-16. https://doi.org/10.20429/ijsotl.2007.010203

Poole, G., 2009. The Pursuit of the Scholarship of Teaching and Learning in Canada: Good, but Not Good Enough. In: *Annual Canadian Society for Studies in Higher Education Conference,* 25-27 May 2009, Ottawa.

Regona, M., Yigitcanlar, T., Xia, B. and Li, R.Y.M., 2022. Opportunities and adoption challenges of AI in the construction industry: a PRISMA review. *Journal of Open Innovation: Technology, Market, and Complexity*, 8(1), 45. https://doi.org/10.3390/joitmc8010045

Rethlefsen, M.L., Kirtley, S., Waffenschmidt, S., Ayala, A.P., Moher, D., Page, M.J. and Koffel, J.B., 2021. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10(1), pp.1-19. https://doi.org/10.1186/s13643-020-01542-z

Rudolph, J., Tan, S. and Tan, S., 2023. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, 6(1). https://doi.org/10.37074/jalt.2023.6.1.9

Rusandi, M.A., Ahman, Saripah, I., Khairun, D.Y. and Mutmainnah, 2023. No worries with ChatGPT: building bridges between artificial intelligence and education with critical thinking soft skills. *Journal of Public Health*, fdad049. https://doi.org/10.1093/pubmed/fdad049

Rushton, A., 2005. ormative assessment: A key to deep learning? *Medical Teacher*, 27(6), pp.509-513. https://doi.org/10.1080/01421590500129159

Rybinski, K. and Kopciuszewska, E., 2021. Will artificial intelligence revolutionise the student evaluation of teaching? A big data study of 1.6 million student reviews. *Assessment & Evaluation in Higher Education*, 46(7), pp.1127-1139. https://doi.org/10.1080/02602938.2020.1844866

Salas-Pilco, S.Z. and Yang, Y., 2022. Artificial intelligence applications in Latin American higher education: a systematic review. *International Journal of Educational Technology in Higher Education*, 19(1), pp.1-20. https://doi.org/10.1186/s41239-022-00326-w

Sarkis-Onofre, R., Catalá-López, F., Aromataris, E. and Lockwood, C., 2021. How to properly use the PRISMA Statement. *Systematic Reviews*, 10(1), pp.1-3. https://doi.org/10.1186/s13643-021-01671-z

Shen, Y., Heacock, L., Elias, J., Hentel, K.D., Reig, B., Shih, G. and Moy, L., 2023. ChatGPT and other large language models are double-edged swords. *Radiology*, 307(2), pp.1-4. https://doi.org/10.1148/radiol.230163
Shiri, A., 2023. ChatGPT and academic integrity. *SSRN Electronic Journal*, 3(2). https://doi.org/10.2139/SSRN.4360052
Spies, A. R., Kjos, A. L., Miesner, A., Chesnut, R., Fink, J. L., D'antonio, N. and Russo-Alvarez, G., 2010. Use of laptops and other technology in the classroom. *American Journal of Pharmaceutical Education*, 74(8). https://doi.org/10.5688/aj7408152
Stokel-Walker, C. 2022. AI bot ChatGPT writes smart essays — should professors worry? *Nature*. https://doi.org/10.1038/d41586-022-04397-7
Susnjak, T., 2022. ChatGPT: The end of online exam integrity? *arXiv*. https://doi.org/10.48550/arXiv.2212.09292
Tibaná-Herrera, G., Fernández-Bajón, M. T. and Moya-Anegón, D., 2018. Categorization of E-learning as an emerging discipline in the world publication system: a bibliometric study in SCOPUS. *International journal of educational technology in higher education*, 15(1), pp.1-23. https://doi.org/10.1186/s41239-018-0103-4
Van Eck, N. J. and Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), pp.523-538. https://doi.org/10.1007/s11192-009-0146-3
Van Haneghan, J., 2021. Exploratory factor analysis of text. In: A.J. Onwuegbuzie and R. Burke Johnson. ed. 2021. *The Routledge reviewer's guide to mixed methods analysis*. London: Routledge, pp. 25-36.
Wang, S., Sun, Z. and Chen, Y., 2023. Effects of higher education institutes' artificial intelligence capability on students' self-efficacy, creativity and learning performance. *Education and Information Technologies*, 28(5), pp.4919-4939. https://doi.org/10.1007/s10639-022-11338-4
Wang, Y., Liu, C. and Tu, Y. F., 2021. Factors affecting the adoption of AI-based applications in higher education. *Educational Technology & Society*, 24(3), pp.116-129. https://doi.org/10.1007/s10639-022-11338-4
Watanabe, A., 2022. Let's Talk About Artificial Intelligence: How Scholarship of Teaching and Learning Can Enhance the AI Scientific Discourse in Higher Education. In: F. Almaraz-Menéndez, A. Maz-Machado, C. López-Esteban and C. Almaraz-López, ed. 2022. *Strategy, Policy, Practice, and Governance for AI in Higher Education Institutions*. IGI Global. pp. 48-72. https://doi.org/10.4018/978-1-7998-9247-2.ch003
Welker, K. M., Duong, M., Rakhshani, A., Dieffenbach, M., Coleman, P. and Haidt, J., 2023. The Online Educational Program 'Perspectives' Improves Affective Polarization, Intellectual Humility, and Conflict Management. *Journal of Social and Political Psychology*, 11(2), pp.437-457. https://doi.org/10.5964/jspp.10651
Williams, A., Verwood, R., Beery, T.A., Dalton, H., McKinnon, J., Strickland, K., Pace, J. and Poole, G., 2013. The Power of Social Networks: A Model for Weaving the Scholarship of Teaching and Learning into Institutional Culture. *Teaching and Learning Inquiry*, 1(2), pp.49-62. https://doi.org/10.2979/teachlearninqu.1.2.49
Yu, L., and Yu, Z., 2023. Qualitative and quantitative analyses of artificial intelligence ethics in education using VOSviewer and CitNetExplorer. *Frontiers in Psychology*, 14, 1061778. https://doi.org/10.3389/fpsyg.2023.1061778
Zawacki-Richter, O., Marín, V., Bond, M. and Gouverneur, F., 2019. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16. https://doi.org/10.1186/s41239-019-0171-0
Zeba, G., Dabić, M., Čičak, M., Daim, T., and Yalcin, H., 2021. Technology mining: Artificial intelligence in manufacturing. *Technological Forecasting and Social Change*, 171, 120971. https://doi.org/10.1016/j.techfore.2021.120971
Zhai, C., 2023. A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence*, 100134. https://doi.org/10.1016/j.caeai.2023.100134

Appendix. Hot articles and hot journals

Table 4: Hot articles. Top 25 by NIY

Article	Authors	Journal	Year	Citations	NIY
Systematic review of research on artificial intelligence applications in higher education - where are the educators?	Zawacki-Richter, O; Marin, VI; Bond, M; Gouverneur, F	International Journal Of Educational Technology In Higher Education	2019	292	58.4
A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education	King, MR	Cellular And Molecular Bioengineering	2023	38	38.0
Adoption of artificial intelligence in higher education: a quantitative analysis using structural equation modelling	Chatterjee, S; Bhattacharjee, KK	Education And Information Technologies	2020	66	16.5
Generative AI and the future of education: Ragnarok or reformation? A paradoxical perspective from management educators	Lim, WM; Gunasekara, A; Pallant, JL; Pallant, JI; Pechenkina, E	International Journal Of Management Education	2023	10	10.0

Article	Authors	Journal	Year	Citations	NIY
A decolonial approach to AI in higher education teaching and learning: strategies for undoing the ethics of digital neocolonialism	Zembylas, M	Learning Media And Technology	2023	9	9.0
Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020	Ouyang, F; Zheng, LY; Jiao, PC	Education And Information Technologies	2022	17	8.5
Prerequisites for artificial intelligence in further education: identification of drivers, barriers, and business models of educational technology companies	Renz, A; Hilbig, R	International Journal Of Educational Technology In Higher Education	2020	34	8.5
Exploring Opportunities and Challenges of Artificial Intelligence and Machine Learning in Higher Education Institutions	Kuleto, V; Ilic, M; Dumangiu, M; Rankovic, M; Martins, OMD; Paun, D; Mihoreanu, L	Sustainability	2021	24	8.0
Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness	Liu, H; Peng, HM; Song, XY; Xu, CZ; Zhang, M	Internet Interventions-The Application Of Information Technology In Mental And Behavioural Health	2022	15	7.5
The intelligent library Thought leaders' views on the likely impact of artificial intelligence on academic libraries	Cox, AM; Pinfield, S; Rutter, S	Library Hi Tech	2019	36	7.2
Blockchain Technology Enhances Sustainable Higher Education	Bucea-Manea-Tonis, R; Martins, OMD; Bucea-Manea-Tonis, R; Gheorghita, C; Kuleto, V; Ilic, MP; Simion, VE	Sustainability	2021	21	7.0
Exploring the impact of Artificial Intelligence and robots on higher education through literature-based design fictions	Cox, AM	International Journal Of Educational Technology In Higher Education	2021	20	6.7
Application of Artificial Intelligence powered digital writing assistant in higher education: randomized controlled trial	Nazari, N; Shabbir, MS; Setiawan, R	Heliyon	2021	19	6.3
Building University-Industry Co-Innovation Networks in Transnational Innovation Ecosystems: Towards a Transdisciplinary Approach of Integrating Social Sciences and Artificial Intelligence	Cai, YZ; Ferrer, BR; Lastra, JLM	Sustainability	2019	31	6.2
Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword	Dahmen, J; Kayaalp, ME; Ollivier, M; Pareek, A; Hirschmann, MT; Karlsson, J; Winkler, PW	Knee Surgery Sports Traumatology Arthroscopy	2023	6	6.0
Future Medical Artificial Intelligence Application Requirements and Expectations of Physicians in German University Hospitals: Web-Based Survey	Maassen, O; Fritsch, S; Palm, J; Deffge, S; Kunze, J; Marx, G; Riedel, M; Schuppert, A; Bickenbach, J	Journal Of Medical Internet Research	2021	18	6.0
The Potential of Blockchain Technology in Higher Education as Perceived by Students in Serbia, Romania, and Portugal	Kuleto, V; Bucea-Manea-Tonis, R; Bucea-Manea-Tonis, R; Ilic, MP; Martins, OMD; Rankovic, M; Coelho, AS	Sustainability	2022	11	5.5
Artificial Intelligence, Chatbots, Plagiarism and Basic Honesty: Comment	Kleebayoon, A; Wiwanitkit, V	Cellular And Molecular Bioengineering	2023	5	5.0

Article	Authors	Journal	Year	Citations	NIY
Remote proctored exams: Integrity assurance in online education?	Paredes, SG; Pena, FDJ; Alcazar, JMD	Distance Education	2021	15	5.0
Can artificial intelligence transform higher education?	Bates, T; Cobo, C; Marino, O; Wheeler, S	International Journal Of Educational Technology In Higher Education	2020	19	4.8
Stress, Coping, and Resilience Before and After COVID-19: A Predictive Model Based on Artificial Intelligence in the University Environment	Morales-Rodriguez, FM; Martinez-Ramon, JP; Mendez, I; Ruiz-Esteban, C	Frontiers In Psychology	2021	14	4.7
Psychometric Properties of the SAS, BAI, and S-AI in Chinese University Students	Pang, ZY; Tu, DB; Cai, Y	Frontiers In Psychology	2019	23	4.6
Artificial intelligence applications in Latin American higher education: a systematic review	Salas-Pilco, SZ; Yang, YQ	International Journal Of Educational Technology In Higher Education	2022	9	4.5
Artificial intelligence (AI) library services innovative conceptual framework for the digital transformation of university education	Okunlaya, RO; Abdullah, NS; Alias, RA	Library Hi Tech	2022	9	4.5
Effectiveness of ideological and political education reform in universities based on data mining artificial intelligence technology	Huang, XY; Zhao, JZ; Fu, JY; Zhang, XX	Journal Of Intelligent & Fuzzy Systems	2021	13	4.3

Note. NIY = Normalized Impact per Year

Table 5: Hot journals. Top 25 by first quartile in Average NIY

Journal	Articles	Citations	NIY (Av.)
Cellular And Molecular Bioengineering	2	43	21.5
Learning Media And Technology	1	9	9.0
Internet Interventions-The Application Of Information Technology In Mental And Behavioural Health	1	15	7.5
International Journal Of Educational Technology In Higher Education	12	383	7.4
Heliyon	1	19	6.3
Journal Of Medical Internet Research	1	18	6.0
Knee Surgery Sports Traumatology Arthroscopy	1	6	6.0
Distance Education	1	15	5.0
International Journal Of Management Education	2	10	5.0
Library Hi Tech	3	51	4.9
Education And Information Technologies	7	86	4.0
Innovations In Education And Teaching International	1	4	4.0
Computer Applications In Engineering Education	1	11	3.7
Electronics	3	24	3.1
Sustainability	14	121	3.1
Journal Of Intellectual Capital	1	12	3.0
British Journal Of Social Work	1	5	2.5
Assessment & Evaluation In Higher Education	1	7	2.3
Applied Sciences-Basel	3	15	2.1
Frontiers In Psychology	6	43	2.0
Biology Of Sport	1	2	2.0

Journal	Articles	Citations	NIY (Av.)
Expert Systems	1	2	2.0
Information Technology & People	1	2	2.0
Mobile Networks & Applications	1	6	2.0
Science Technology & Human Values	1	14	2.0

Note. Av.NIY = Average normalized impact per year for all articles published by each journal

Exploring Student and AI Generated Texts: Reflections on Reflection Texts

Marcia Håkansson Lindqvist and Catarina Arvidsson
Department of Education, Mid Sweden University, Sundsvall, Sweden

marcia.hakanssonlindqvist@miun.se (Corresponding author)
catarina.arvidsson@miun.se

https://doi.org/10.34190/ejel.22.6.3473

Abstract: As pointed out by many scholars, Artificial Intelligence (AI) provides both opportunities and challenges in regard to assignments and examination in higher education. The accessibility and use of AI in regard to student assignments, examinations and assessments places demands on teachers' work in course design and formats of assignments and examination. For teachers, this work is a constant and continuous process, in line with the Scholarship of Teaching and Learning (SoTL) according to Boyer (1991). In order to meet these new demands, teachers need to reflect upon design, as reflective practitioners (Schön, 1987). Reflective design may alleviate the challenges with AI as well as make use of the opportunities with the use of AI. In this paper there are two sets of data. This study aspires to contribute to the current state of AI (ChatGPT) as it is applied in higher education through an empirical study of authentic reflection texts by students in comparison to AI (ChatGPT) generated texts. The first set of data is authentic reflection texts (N=20) written by students. The second set of data is texts generated by AI (ChatGPT). The texts are analysed using reflective thematic analysis (Braun & Clarke, 2019). The themes in the two sets of texts are described, analysed and compared. The two sets of data are then explored, analysed and compared to highlight similarities and differences between the authentic texts and the texts generated by AI. These insights may provide support for teachers in regard to the design of assignments and examinations as well as the practical use of AI (ChatGPT) in higher education.

Keywords: Assignments, Examination, Higher education, Reflection, Teachers

1. Introduction

The many uses of Artificial Intelligence (AI) have been in focus as of late. This has been especially true in the area of higher education, where teachers debate the possibilities and challenges in using AI in education to support student learning. The issue, for example, has involved to use or not to use AI (ChatGPT) in education (Strzelecki, 2023). Policies and easy-to- understand guidelines for the use of language models in learning and teaching have been put in the forefront in order to support proper use of these tools and to limit the consequences for cheating (Rudolph, Tan & Tan, 2023). Neumann, Rauschenberger and Schön (2023) report five challenges and three opportunities for the use of AI in education. Challenges involve not knowing how students employ AI, different ways in which teachers evaluate AI, ideas on what is considered to be acceptable or unacceptable use, more time-consuming assessments and the uncertainty regarding the unknown potential of AI. However, Neumann, Rauschenberger and Schön (2023) also see possibilities such as increased virtual support for students, engaged creativity for teachers and students and potential for innovation.

Nevertheless, there is the need for teachers to critically evaluate AI resources and adapt it to their specific context. This places the focus on the importance of teacher's expertise, experience, and understanding of their students (Cooper, 2023) in order to support the use of AI. Thus, it is most likely that training in these tools is necessary, i. e. there is a need to conduct training for faculty and to make students aware of academic integrity (Rudolph, Tan & Tan, 2023). Thus, there is a need for further research on how AI tools can be adopted in teaching and learning (Strzelecki, 2023). In reflection, many of these studies provide important insights into the challenges of using AI in education, while some note possibilities. Nevertheless, beyond recommendations for training and the need to acknowledge critical issues, few studies delve deeper in the use of AI in practice in teaching and learning activities to support teachers' work with AI with students.

In higher education, the use of critical reflection is of importance for teachers as reflective practitioners (Bie, 2014; Schön, 1987). The importance of practicing reflection is necessary in order to develop in one's role as an educator, as an individual and as a member of a work group or group of colleagues, according to Bie (2014). Moreover, the use of reflection as learning tool is important to so see oneself (Bie, 2014). The idea of reflection as a tool for learning (Bie, 2014; Schön, 1987) is also important for teachers and students in order to continually

Cite this article: Håkansson Lindqvist, M. and Arvidsson, C. 2024. "Exploring Student and AI Generated Texts: Reflections on Reflection Texts", *Electronic Journal of e-Learning*, 22(6), pp 52-59, https://doi.org/10.34190/ejel.22.6.3473

reflect upon in order to develop practice. The constant strife to develop teaching and learning is described as an exploratory approach to the role of teacher in line with Scholarship of Teaching and Learning (SoTL) according to Boyer (1991). As a reflective practitioner, the use of reflection as a tool for learning can also be seen as a tool for development of teaching and learning practices. However, for students, reflection as a tool for learning is sometimes difficult (Bie, 2014). Students may need support in learning to use reflection as a tool for learning as well as to achieve higher levels of reflection (Bie, 2014).

Hatton and Smith (1995) describe four levels of reflection: *Descriptive writing, descriptive reflection, dialogic reflection* and *critical reflection*. On the first level, descriptive writing, events and behaviour in a situation are described. This level involves mainly description and does not involve reflection. On the second level, descriptive reflection, students describe, based on personal opinions, or write how they think, act and explain the reasons behind the behaviour. This level could be said to involve attempts to reflect. On the third level, dialogic reflection, there is an inner reflective dialogue, in which underlying causes as well as alternative solutions and possible consequences are reflected upon and discussed. On the fourth level, critical reflection, reflection involves reflection from a broader perspective, for example, interaction with other people, or from a societal perspective. The gaze is lifted and the issue is reflected upon in a larger context.

In the light of this short background the notion of reflection as a tool for learning and the study of students' reflection texts and AI (ChatGPT) generated text is in focus in order to gain a wider understanding of the opportunities which AI (ChatGPT) can provide in teaching and learning in practice. The aim of this study is to explore, analyse and compare authentic student reflection texts with AI (ChatGPT) generated texts based on an assigned reflection task. The following research questions were posed: *1) How can the content of the student texts be described? 2) How can the content of AI (ChatGPT) generated texts be described? and 3) Which similarities and differences can be seen?* These insights may provide support for teachers in regard to the design of assignments and examination as well as the practical use of AI (ChatGPT) in higher education.

2. Method

In order to meet this aim and these research questions, authentic reflection texts written by students (N=20) were analysed. These texts were written during a course for behavioral science students studying a basic education course during 2017. The students were asked to write a reflection text based on the following assignment using the course literature on reflection for pedagogues, Bie (2014). Bie (2014) is a handbook for pedagogues which presents reflections on reflection as a tool for learning in different educational contexts. The reflection assignment was:

> *This task is about reflecting on your own learning process during this stage, based on Bie (2014). Some issues to reflect on can be your reading and processing of the literature, your fellow students as resources for learning in your learning process and also reflections on what you will be able to use your new knowledge in future studies and work.*

The texts were short texts, at the longest some 1,000 words. The students were informed about the analysis of the texts and the use for teaching and learning purposes. The students replied individually by e-mail, giving their approval for the use of the texts for the use in the study. The students' texts were analysed using thematic analysis (Braun & Clarke, 2012). Thematic analysis was seen to provide the possibility to analyse the texts through reading and re-reading as a reflective process. In this process, themes were derived and coded. The analysis was done first individually by the authors and the discussed together. In the student texts, the following four themes emerged: *Reflection as a tool for learning, Study groups (fellow students) a resource for learning, Content and course literature* and *Continued studies and future work*. In the analysis of the AI(ChatGPT) texts six themes emerged: *Introduction, Reading and processing the literature, Fellow students as a resource for learning, Use of new knowledge in future studies and work, General reflections on the learning process* and *Concrete suggestions.* In the analysis, the themes that emerged in the texts are presented and exemplified using quotations from the text. The student texts are identified as Text 1-Text 20 (T1-T20).

The same reflection assignment was then put forth in AI (ChatGPT). This process was done five times. Thereafter, the AI generated reflection texts were also analysed using thematic analysis (Braun & Clarke, 2012) in the same manner as the students' texts were analysed. The AI (ChatGPT) generated texts (N=5) are identified as AI1-AI5. Thereafter, the themes which emerged in the authentic student texts were compared with responses to the same question from AI (ChatGPT) generated texts.

3. Results

In this section the results are presented according to the themes which emerged in the analysis. First, the student texts are presented, and thereafter the AI (ChatGPT) generated texts are presented.

3.1 Authentic Student Texts

In the student texts, the following themes were seen Reflection as a tool for learning, Study groups (fellow students) a resource for learning, Content and course literature and Continued studies and future work.

In the first theme, Reflection as a tool for learning, students wrote about study technique and study methods. The students also reflected on individual insights and reflections on how reflection can support learning. In this theme, students also reflected on own activities which they found to be important to support their own learning. One student reflected upon new study techniques:

> *While reading, I try to "read with pen in hand", which is a piece of advice we received early in the course. It's often easier to remember and understand the most important and central parts of the course literature when I do this, but it sometimes feels difficult when I later try to understand my own notes. I need to work a bit on this technique, because I realize that it is probably the best for me as long as I get a bit more practice with it. (T11)*

Another student pointed out the importance of slowing down and taking time for reflection:

> *Bie describes how it is important to reflect after an action and think through something that has happened. In my learning process I have realised how important it is to give myself time to reflect on what I have read to avoid "information cramming" (T2).*

In the second theme, students reflected upon *Study groups (fellow students) as a resource for learning*. This involved how students themselves contributed and took part in learning through the study group and with fellow students. Students also pointed out specific activities in the study group which they considered to support own learning and learning in the group. One student explained:

> *Now the study group feels like one of the most important parts of the programme for me and I think we have reflected well together, maintained good ethical competence and thus increased the security of the group. We have begun to work more effectively together, and I really feel that I get a breadth in what I learn when I can read what my fellow students write and when we discuss the literature and different opinions together. (T8)*

Another student reflected upon working together in the study group in order to create a common understanding:

> *My fellow students are incredibly important resources for my own learning and we talk a few times a week when we discuss different literature and what we have come up with. Sometimes we don't understand and then we can sort out and understand the meaning of the literature together. (T5)*

In the third theme students reflected upon the *Content and course literature*: This involved taking on the literature, course design and evaluations of the course literature. One student reflected on taking on the course literature:

> *Thanks to the generous amount of time set aside for reading, I have had the pleasure of reading each chapter as thoroughly as I wish without having to prioritize and jump around the texts according to estimated relevance. I appreciate this because it gives me a better sense of coherence and overall picture of the literature (T14).*

By developing own study techniques, it was easier to have a comprehensive view of the course and the course literature. Progress in working with the literature was also noted by one student:

> *I find this stage very rewarding and I notice how the processing of the literature goes more and more smoothly as time goes by. I have developed a working strategy for processing the literature that works well for me. It consists of taking notes while reading. I then use this when writing assignments and exams. Then I just go through the various documents on the computer to get a quick overview of what the literature is about. (T12).*

In this reflection, this student expresses and interesting aspect regarding own negotiation concerning own resistance concerning reading and reflection and the process in the transition in understanding, from sometimes very clear and sometimes quite obscure.

The fourth theme, *Continued studies and future work*, comprised reflections concerning how students reflected on how to use the new knowledge gained in future work. This also involved how new knowledge in continued studies as well as the importance of the new knowledge for studies and work:

> *What I take with me as a future behavioral scientist when it comes to assessment, selection and recruitment is that self-awareness, knowledge of how knowledge and competences can be assessed but also how I can use this knowledge in practice in, for example, assessment and selection in a recruitment process. Knowledge, understanding in the main field of pedagogy and self-awareness are important to be able to make fair and reliable assessments. Another understanding that the course component has contributed to is the importance of lifelong learning in today's knowledge society and how individuals must increasingly take responsibility for their own employability. (T13)*

Another student reflected upon gaining a holistic view of processes and concepts that may be of importance in future studies and work through collaboration:

> *After this part, I have reflected on the importance of a thorough and well-designed recruitment process when a company is looking for new employees. I have also learnt about the concept of employability and its importance in today's society. Without any major experience in the field, I now have a completely different knowledge about assessment, recruitment and selection. I have also reflected on the importance of co-operation and reflective discussions and how instructive it is. Furthermore, I believe that the knowledge I have acquired during the module can be very useful for future studies/work (T6).*

In summary, the reflection assignments written by the students provided in-depth insights into new knowledge regarding their learning in reflection. Here, students reflected upon new learning for themselves as individuals as well as in their study groups and fellow students. Students also reflected on how the course content and literature had been beneficial in their own learning and how the new knowledge they had attained would have impact on their learning in their future studies and work.

3.2 AI (ChatGPT) Generated Texts

In this section, the analysis of the AI (ChatGPT) generated texts is presented. When the reflection assignment was posed to AI (ChatGPT), the first part of the text was an introduction to the reflection assignment and six themes emerged. Here, the themes included: *Introduction, Reading and processing the literature, Fellow students as a resource for learning, Use of new knowledge in future studies and work, General reflections on the learning process* and *Concrete suggestions.*

In the first theme, *Introduction*, the text included information as well as apology for not being able to provide specific answers in regard to course literature at hand, Bie (2014). In this sense, AI (ChatGPT) replied to the reflection assignment in line with answering that *unfortunately, I cannot help you with answers about specific literature* in different in several different ways:

> *Unfortunately, I do not have any specific information on Bie (2014) or details on the particular sub-section you are referring to (AI1).*
>
> *I apologize for any confusion, but I have no specific information on Bie (2014). (AI2).*

One interesting finding was that as the reflection assignment was posed to ChatGPT several times, what could be said to be impatience was noted in in the text:

> *I still have no information about Bie (2014), but I can help you create reflection questions based on the general aspects of learning. (AI5)*

Another finding was that ChatGPT offered several different methods of possibilities for further help, for example "I can help you with several reflection questions to help you work further .... general guidelines" (A11) as well as "general questions" (AI3) "questions" (AI4) and "some questions" (AI5). In summary, the first part of the texts could be said to be a limitation of the possibilities to provide specific information regarding the specific course literature as well as providing questions to support further work.

In the second theme, *Reading and processing the literature,* the responses in the AI (ChatGPT) generated text provided reflection questions which supported reading and processing the course literature:

> *How did you structure your reading of the literature? Did you use any special strategies or techniques? What challenges did you encounter during your reading, and how did you overcome them? (AI1)*

These reflection questions referred to the methods used to read and process the literature as well key concepts and theories in relation to understanding and learning through the work with the literature. In these questions the reference to the literature was included:

> *What method did you use to read and process the material from Bie (2014)? Are there any key concepts or theories from the text that influenced your understanding of the topic? (AI3)*

> *Which main ideas or concepts from Bie (2014) were particularly relevant for this sub-step (AI5)?*

The third theme was related to *Fellow students as a resource for learning*. Here, the texts provided responses which provided reflection questions for identifying and describing how fellow students were seen as a resource for learning. These responses provoked further reflection into methods of collaboration and interaction with fellow students as a resource or tool for learning and understanding:

> *How did you use your fellow students as a resource for learning? Were there any specific collaboration methods that worked well? (AI1)*

> *How has interaction with fellow students affected your understanding of the topic? (AI2)*

Other questions for support in reflection were providing specific examples which provided students with learning opportunities:

> *Can you point to any specific examples where you have learned from your peers or vice versa? (AI5)*

In the fourth theme, *Use of new knowledge in future studies and work,* the AI generated texts provided responses with reflection questions regarding new knowledge, skills or insights which could be seen as important for future studies.

> *How do you think the new knowledge you have acquired will be useful in your future studies? (AI1)*

> *Can you identify specific skills or insights that you can apply in your future career? Are there any areas where you see opportunities to continue learning and developing your skills? (AI1)*

This also involved skills development for the future profession. Here, the AI generated texts and reflection questions pointed out the necessity to first identify and then reflect upon how these skills could be of use and of importance in future work:

> *Can you identify specific skills or insights from this module that you think will be useful in your future work? (AI4)*

This theme also included more individual, or personal reflections on own learning. For example, these reflection questions noted the identification and further work with reflection as a tool for continued learning and development:

> *What were the most meaningful parts of the learning process for you personally? Are there any aspects of the studies that you would like to improve or deepen further? (AI1)*

The fifth theme which emerged was *General reflections* on the learning process. In this theme the reflection questions in the AI (ChatGPT) generated texts were focused on reflection upon learning and changes in learning due to new knowledge attained in the learning process:

> *Have your views on the topic changed or evolved during this stage? Is there anything you would do differently if you had the opportunity to go back and re-evaluate your approach? (AI4)*

> *What were the most meaningful parts of the learning process for you personally? Are there any aspects of the studies that you would like to improve or deepen further? (AI1)*

> *Have your views on the topic changed or evolved during this stage? Is there anything you would do differently if you had the opportunity to go back and re-evaluate your approach? (AI4)*

In summary, these reflection questions provided a base for meta-reflection regarding learning, learning processes and evaluation of these learning processes.

The sixth and final theme *Concrete examples* referred to reflection questions which asked specifically for concrete examples to provide in the reflection assignment. In the AI (ChatGPT) generated texts, concrete examples, or specific examples were suggested to provided examples of learning processes and new knowledge:

> *Try to be as concrete as possible when reflecting on these questions. Link your answers to specific examples or events during the learning process to make your reflection more meaningful and useful for your further development. (AI2)*

> *Use these questions as a basis and adapt them to your own experiences and learnings from the specific module. It is important to be as concrete as possible to gain deeper insights and use the reflection as a tool for personal and professional development (AI5). This theme could be said to connect reflection as a tool for learning in theory with concrete examples in practice as a base for reflection.*

In summary, the results of this study show that the content in the authentic student texts can be described as individual, personal, reflective texts related to study methods and study techniques for learning used both individually and in study groups with fellow students. Reflection as a tool for learning and knowledge is then extended and identified as a tool for learning in future studies and work. The content in the AI (ChatGPT) generated texts, which perhaps could be considered to be somewhat surprising, did not provide individual or personal reflections related to reflection as a tool for learning and new knowledge, but instead provided a battery of supportive and in-depth reflection questions as inspiration for writing the reflection assignment.

4. Discussion

The aim of this study was to explore, analyse and compare authentic student reflection texts with AI (ChatGPT)generated texts based on an assigned reflection task. The following research questions were posed: 1) How can the content of the student texts be described? 2) How can the content of AI (ChatGPT) generated texts be described? and 3) Which similarities and differences can be seen?

The findings in this study show that the AI (ChatGPT) generated texts provided reflection questions which are of an in-depth nature. Thus, in this study the use of AI (ChatGPT) generated texts did not provided students with answers to the assignment which could be submitted for grading. Therefore, in this study, the realization of the use of AI (ChatGPT) could not be seen as a challenge in regard to cheating (Neumann, Rauschenberger & Schön, 2023; Rudolph, Tan & Tan, 2023). On the contrary, the AI (ChatGPT) responses to the reflection assignment in this study could instead be seen as an opportunity to support students' further work with the assignment.

Reflection as a tool for learning is not an easy task (Bie, 2014). Many students may have difficulties in writing reflection texts as well as achieving higher levels of reflection in reflection texts (Hatton & Smith, 1995). Here, it is possible that the reflection questions are concrete and can help students to expand and advance their use of reflection as a tool for learning. Further, the in-depth reflection questions are directed to the students themselves. The reflection questions target the identification and description of specific or concrete examples as evidence for learning through reflection. This can provide support for students when reviewing their own learning, reflecting upon their own study strategies and providing inspiration for further reflection regarding learning, both individually and in study groups with fellow students or peers.

In this study, reflection as a tool for learning, was explored related to future studies and work. In reflection, work with students could also involve important ethical implications regarding the use of AI educational contexts. Teachers' work with AI together with students could be directed toward other important issues such as academic integrity, privacy issues and data security. These issues could also be involved in tailored or model assignments for use in practice, for example for student teachers in their work with pupils in different education contexts. Here, both the teacher student perspective and the pupil perspective could provide insights from the tools from the learner perspective. These aspects would provide opportunities for teachers and students to both learn *about* AI and *with* AI to deepen teaching and learning, offering valuable insights in theory and practice.

As teachers in higher education, work in line with reflection as a tool for learning in teaching in higher education, the findings of this study warrant further reflection. As noted above, in this study, the AI (ChatGPT) generated texts did not provide students with texts that they could submit. Therefore, as teachers, concerns and apprehension regarding the use of AI (ChatGPT) as a tool for cheating were not relevant. Instead, the AI (ChatGPT) generated texts provide further support for students to achieve deeper reflection. This line of thought leads to the analysis of the critical levels of reflection in regard to AI and student generated texts. Reflective analysis may perhaps offer deeper insights in how AI can or cannot mimic complex human cognitive processes, such as deep levels of reflection and critical thinking. As reflective practitioners, one reflection is that these in-depth reflection questions may support students in their individual work but would most likely provide even more support if used under the supervision of teachers. In regard to accessibility and inclusion, this may especially be true for students who lack the technical infrastructure for these tools. For students with disabilities, the integration of AI may provide extended opportunities for enhanced learning. Moreover, given the importance of AI in technology enhanced learning environments and e-learning environments, the findings of this study may also be applied to enhance learning in remote and online learning environments. Furthermore, used well, AI (ChatGPT) may support teachers' work through the formulation of assignments (Cooper, 2023). For example, it is possible that some the reflection questions generated by AI (ChatGPT) could be integrated into the reflection assignment in the future. However, this work will involve pedagogical discussion in teacher groups regarding the challenges and opportunities of AI in education (Neumann, Rauschenberger & Schön, 2023; Strzelecki, 2023). Beyond the challenges and opportunities for teaching and learning which AI provides at present, future technologies and advancement in AI will most likely have strong impact on educational contexts. This strong impact may have the potential to transform teaching and learning and therefore offering exciting and expansive venues for future research.

5. Future Research

One of the limitations in this study is that the sample of both authentic student texts and AI (ChatGPT) generated texts is small. Thus, this study could be expanded to a larger sample and the study of different types of reflection questions. Continued work with AI (ChatGPT), as well as with other AI models, to support reflection as a tool for learning for students should involve students in order to gain important insights from the learner perspective. Case studies which involve students in the comparative analysis regarding the effectiveness of these different tools, could also provide insights from students as learners regarding how the current technologies in use enable and hinder collaborative learning. Therefore, future studies should involve students in questions concerning how AI (ChatGPT) can be used to support student learning as well as the study of what challenges and opportunities students see in this work. Practical applications in practice could take issues on academic integrity, privacy issues and data security where teachers integrate AI in teaching and learning activities and employ AI for deeper reflective questioning and discussion. Larger empirical studies which involve teachers in higher education, related to assignments and assessment and the use of AI (ChatGPT) would provide insights into the challenges, as well more balanced insights in the opportunities for the use of AI (ChatGPT) in higher education. Further, interdisciplinary studies involving cognitive science, ethics and technology enhanced learning could also provide a wider, richer and more nuanced analysis of the implications in practice and in theory of the use of AI in education to support teaching and learning.

Funding

No funding was received for the work with this article.

Conflict of interest

The authors declare no conflict of interest.

Ethics statement

An ethics approval was not required for this study.

Acknowledgements

We would like to acknowledge the students in this study, whose texts made this study possible.

References

Bie, K. (2014). *Reflektionsboken för pedagoger* [Reflection book for pedagogues]. Gleerups.

Boyer, E. (1991). Highlights of the Carnegie report. The Scholarship of Teaching from "Scholarship reconsidered: Priorities of the Professoriate". *College Teaching, 39*(1), 11-13.

Braun, V. & Clarke, V. (2012). Thematic analysis. I: H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Red.), *APA handbook of research methods in psychology, vol. 2: Research designs: Quantitative, qualitative, neuropsychological, and biological* (s. 57–71). American Psychological Association.

Cooper, G. (2023). Examining science education in chatgpt: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology, 32*(3), 444-452.

Hatton, S. & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching & Teacher Education, 11*(1), 33-49.

Neumann, M., Rauschenberger, M., & Schön, E. M. (2023). "We Need To Talk About ChatGPT": *The Future of AI and Higher Education. Hochschule Hannover, 2002*(1), 1–4.

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1).

Schön, D. (1987). *Educating the reflective practitioner. Toward a new design for teaching and learning in the profession*. Jossey Bass.

Strzelecki, A. (2023). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments*, 1-14.

Risks of AI Applications Used in Higher Education

Donna Schaeffer[1], Lori Coombs[1], Jonathan Luckett[1], Marvin Marin[1] and Patrick Olson[2]
[1]Marymount University, School of Technology and Innovation, Arlington, USA
[2]National University, San Diego, USA

Donna.schaeffer@marymount.edu (corresponding author)
ldc72459@marymount.edu
jpl65485@marymount.edu
m0m61652@marymount.edu
polson@nu.edu

https://doi.org/10.34190/ejel.22.6.3457

Abstract: As artificial intelligence (AI) tools become more widely used in higher education, we must pay attention to the risks that can emerge. AI projects, whether applied in classroom learning or used for decision-making regarding admissions, financial aid allocation, or hiring, must include attention to governance and compliance issues, regardless of the project's scope and scale. Concerns highlighted in this work include transparency, user privacy, data confidentiality, data integrity, and system availability, however, we note that this is a non-exhaustive list of risks. In this paper, risk assessment is defined, and two examples of risk management frameworks, namely the United States National Institute of Standards and Technology Artificial Intelligence Risk Management Framework and the non-profit humanitarian effort ForHumanity's Independent Audit of AI, Algorithmic, and Autonomous Systems are briefly described. We identify characteristics of AI applications that need to be assessed for vulnerabilities they may present, such as bias and discrimination. This paper aims to facilitate discussion among stakeholders about the risks that may be encountered from using AI in higher education, as well as to suggest ways developers, decision-makers, and users can mitigate these risks. Much discussion and published literature has focused on risk management frameworks designed for large organizations or enterprises or frameworks that do not consider risks specific to AI. We hope that decision-makers carefully consider the risks, perform due diligence when implementing AI applications, and create a plan for mitigating the risks. This research supports e-learning practice because students and faculty are embracing AI applications. Leaders and decision-makers in higher education need to be proactive in protecting their varied stakeholders. The paper asks what risks may be encountered by institutions of higher education when using AI tools and products in the classroom and for various aspects of decision-making and if published frameworks can mitigate these risks.

Keywords: Artificial intelligence (AI), Risk management framework (RMF), Higher education, Cybersecurity

1. Introduction

This paper describes the risks that higher education institutions face as the use of artificial intelligence (AI) is becoming prevalent. This paper presents definitions, describes risk management frameworks, and discusses implications that decision-makers must consider.

The current discourse perceives AI as a two-edged sword when considering its impact on academia. The first edge of the sword questions ethical use. For example, AI Large Language Model (LLM) systems, such as ChatGPT, empower students with the ability to create content on behalf of prompts, which is known as prompt engineering, and those ideas can be used in whole or in part by students to submit as their work. At the very least, this creates a question of authorship. Does the student get the credit for generating the idea, or does the credit go to the AI for putting the words in the order that may be edited by students? Additionally, what is the demarcation line? If a writer uses a software tool such as Microsoft Word, and that tool has a rudimentary predictive element to it, does the credit belong to the human or the AI component – the predictive tool?

The second edge of the sword is AI's potential as a game-changer for students with disabilities that impact their academic performance. For example, students who suffer from dyscalculia may have an issue with understanding or conceptualizing mathematical concepts. However, a university's accommodations department may work with faculty on a solution where students can access an AI application that can provide additional learning resources. It may have the ability to explain concepts in several ways to reinforce the material, and it can facilitate practicing and reinforcing the lessons learned on the students' schedules when they are most comfortable with learning the material and in a manner that does not fatigue a human educator. In this use, students are not being given an advantage; instead, they are being allowed to be as successful as neurotypical students. This level of equity can be achieved by making AI available to students within this population. What is

Cite this article: Schaeffer, D.et al. 2024. "Risks of AI Applications Used in Higher Education", *Electronic Journal of e-Learning*, 22(6), pp 60-65, https://doi.org/10.34190/ejel.22.6.3457

important to the university and faculty is that students can understand the material presented and, with accommodation(s), convey competence in the material taught following the learning objectives in the course syllabus.

This discussion reveals a unique relationship the education sector has with AI. On the one hand, AI as a field within computer science should be explored for its ability to solve complex problems quickly and redefine how we educate students. On the other hand, it can be a major disruptor in how students' work is created and assessed, how educators approach education and the faculty-student relationship, and how decision-makers allocate resources. The conversation in higher education must shift to AI as a means of learning and teaching the next generation. These are the tools that the current generation of students will invariably use in the future; academics cannot refuse to teach about these valuable tools. Ignoring AI or making AI applications unavailable to students will not put them in a position to excel in the careers of the future and to create new knowledge. However, we must instill within students an ethical understanding and application for the use of AI. In many ways, this is no different from the lessons learned in business schools and the introduction of ethics classes in the curriculum of Master of Business Administration (MBA) degree programs after the collapse of Enron and other economic failures. Embracing AI is practical and prudent. Academia should embrace that students may use AI unethically and develop guidelines, policies, regulations, and educational material to help students understand what is required of them and how they can use the tools available to them. Given that many universities already educate their student population on topics such as academic integrity and academic misconduct, educating them about the ethical use of AI would be a *minimal* expansion of existing efforts.

While the most visible applications of AI may be in curriculum and pedagogy, those applications that touch students directly, much use of AI will happen "behind the scenes" in recruiting and admissions scenarios, allocating resources, and planning. This use will be by university and college administrators and staff and may even be outsourced to third-party vendors. Administrators must have a risk management strategy for outcomes these applications may produce. Risk management strategies typically are built around a risk management framework. After defining key terms and risk scenarios, this paper will describe two risk management frameworks.

2. Definitions

This paper standardizes on definitions provided by the National Institute of Standards and Technology (NIST), an organization within the United States Department of Commerce. Dempsey et al. (2011) define risk as “A measure of the extent to which an entity is threatened by a potential circumstance or event, and typically a function of: (i) the adverse impacts that would arise if the circumstance or event occurs; and (ii) the likelihood of occurrence (Page B-10).” This is often stated as the equation Risk = threats x vulnerabilities.

As NIST becomes involved in the standardization of AI across the US federal government, it has chosen to follow the American Standard Dictionary of Information Technology definition of AI, that is (1) A branch of computer science devoted to developing data processing systems that perform functions commonly associated with human intelligence, such as reasoning, learning, and self-improvement and (2) The capability of a device to perform functions that are normally associated with human intelligence such as reasoning, learning, and self-improvement (National Institute of Standards and Technology, 2019, page 25.). AI is a complex field that encompasses diverse technologies. For example, rule-based AI applications leverage methods that program computers to make decisions based on a set of rules. In contrast, expert systems are programmed to emulate the decision-making abilities of human experts. Machine Language (ML) is a subset of AI that leverages algorithms, data sets, and models to perform specific tasks. Deep learning solves problems in the same manner as the human brain does by using algorithms; however, it requires more extensive data sets than ML. Generative AI is an emerging development that has raised concern in governments and the general public for intellectual property infringement because the AI application draws on patterns that often utilize data lakes (centralized repositories that allow structured and unstructured data *to* be stored) and unlicensed work (Appel et al., 2023).

Several frameworks define risk management in AI. A Risk Management Framework (RMF) is a structured approach used to oversee and manage risk for an enterprise (Nieles, et al., 2017). This paper discusses two RMFs – The National Institute of Science and Technology (NIST) AI Risk Management Framework (AIRMF), which incorporates international standards such as ISO/IEC 23984:2023, and the non-profit organization ForHumanity, which has developed a certification program for auditors around the ethical assessment of AI applications.

This paper's scope of risks includes transparency, privacy, confidentiality, data integrity, system availability, and bias. The reasons these risks are important to consider are explained within the definitions provided in this section of the paper. While the NIST definition of transparency is narrow – its publications define transparency as the amount of information that can be gathered about a supplier, product, or service and how far through the supply chain this information can be obtained (Boyens et al., 2021), it also offers transparency as a synonym for visibility. In AI applications, transparency can be broadened to denote how much can be understood about the AI application, including how it was developed and trained and how it operates when deployed. Users must have access to the data sets that were used to train the model. The characteristic of transparency can illuminate issues of bias. If users understand why the AI application results in the recommendations it provides, they can decide whether to accept or reject the output. To mitigate problems associated with security, trust, and objectivity, models should integrate transparency when designing algorithms.

Any discussion of transparency yields decision points. Systems engineers, developers, and systems administrators who develop and deploy AI in educational institutions should consider engaging with lawyers early in the design process. AI and ML algorithms need to be examined for vulnerabilities and liabilities from cybersecurity and human user experience perspectives. Higher education institutions need enhanced transparency for AI models and machine learning algorithms. However, developers may not want to expose their code, patterns, and data sets to users because that level of openness could make applications vulnerable to attacks. While generating more information about the AI application might create tangible benefits for users, it may also create new risks for developers and users (Burt, 2019).

Privacy and confidentiality are related concepts. Indeed, NIST uses the terms in one another's definitions. For example, Powell et al. (2022) define privacy as the assurance that the confidentiality of and access to certain information about an entity is protected. Confidentiality is defined as "Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information (Pub., FIPS, 2006, page 6)." NIST also places privacy in the context of rights – Oldehoeft (1992) states privacy is the right of a party to maintain control over and confidentiality of information about itself. When data is misused, it becomes a source of liability. The data sets that are used to train AI systems must be very large and may include sensitive data. Data sets used for training must not contain personally identifiable information or information that can be aggregated to identify individuals. For example, a recent breach of an AI dataset occurred when developers at Microsoft caused the exposure of 38 terabytes of data, including disk backups of two employees' workstations, confidential corporate information, private keys, passwords, and over 30,000 internal Microsoft Teams messages during a routine update to GitHub (Naraine, 2023). While this example does not involve higher education, it shows the risks of using AI applications**.** Developers must build privacy protection into their systems and applications early in the design phase to protect users. Users should be informed when and how their data is collected and used. When data is collected, users can be offered informed consent agreements, opt-out ability, and the ability to delete their data. AI users must have a method for managing their privacy risks.

Data confidentiality involves protecting information in a system or application so that unauthorized access is prevented. Stakeholders within higher education include administrators, faculty and staff, students, and other related entities. These stakeholders do not want their sensitive and personally identifiable information (PII) exposed when institutions use third-party AI applications. Technical staff must understand their shared responsibility involved with combating commonly encountered threats to information confidentiality, including hackers, unprotected downloaded files, unauthorized user activity, local area networks (LANs), and trojan horses. Confidentiality can be compromised in data, network, end-to-end, application, and disk file scenarios. Third-party AI applications should be listed on an approved list for the institution so that service level agreements (SLAs) are put in place that clearly define the university's and the third-party application developer's shared model of responsibility. A constructive shared responsibility model considers access control, encryption, data masking, secure file transfer protocols (SFTPs), data loss prevention (DLP), and virtual private networks (VPNs) to protect the confidentiality of users (Anonymous, 2023).

Data integrity is a property where data or information has not been altered or destroyed in an unauthorized manner (Scholl et al., 2008). Data integrity is rooted in trusted data. Without trusted data from connected sensors, devices, systems, and applications, they all become vulnerable to improper cyber manipulation, making AI decision-making questionable. Data integrity is one pillar of the widely used concept of the "CIA Triad," which highlights confidentiality, integrity, and availability. Any data integrity breach means that AI and related devices won't be able to operate properly, exposing systems and applications to exploitation and cyber-attacks (Armilis, 2023).

Scholl et al. (2008) define system availability as the assurance that users have timely and reliable access to and use of information. Systems availability risk occurs when decisions are made based on easily and immediately available data without considering further research or additional external perspectives. The danger with availability risks when using AI applications is that generative AI is only sometimes accurate or reliable (Pavlou, 2023). When AI yields inaccurate or incomplete results, broader risk implications are present in higher education.

The final risk in this paper's scope is bias. Barker and Kelsey (2007) state that bias exists if one value from a sample space is more likely to be chosen than another value. Bias is discussed in an upcoming section.

These baseline definitions allow us to recognize risks inherent to and introduced through AI application development and deployment. Thus, multiple stakeholders, including developers, administrators in higher education institutions, faculty, students, and other persons who interact with the application, are involved. The definitions also facilitate the identification of risks that may arise in the development and use of AI applications.

3. Underlying Factors That can Cause Risks

Each of the risks defined in the previous section of the paper can be traced back to one of several occurrences that may be inherent in AI and machine learning (ML). These occurrences are data persistence, data repurposing, and data spillover. Pearce (2021) states that data persistence occurs when data exists longer than the developers intend. Data repurposing is when data is used beyond its originally designated purpose (Pearce, 2021). Data spillover occurs when data is collected on entities that are not the intended target of data collection (Pearce, 2021). It is easy to see how any of these occurrences can be present in AI applications because the training sets are often supplemented with additional data as time progresses to fine-tune the system.

Mitigating risk is essential when developing and deploying AI. The Definitions section of this paper includes inherent dangers that become exposed when managing different types of risks. This section briefly describes two Risk Management Frameworks: ForHumanity is designed for developers, and the NIST AI RMF is appropriate for large-scale government enterprises. At the time of this writing, no frameworks consider specific risks for institutions of higher education. The section concludes with a short discussion about how each of these frameworks handles one risk—that of bias.

ForHumanity is a non-profit organization founded in 2016 to develop an independent certification for those who audit AI systems. It is a volunteer-based organization with a small board and close to 1400 volunteers from 89 countries. ForHumanity offers several certifications, such as the ForHumanity Certified Auditor (FHCA), which includes a Code of Ethics. This group has codified a body of knowledge on various areas of compliance that auditors need to consider. Some areas include accuracy, validity, reliability, resilience and robustness, anti-discrimination, and data security. The non-profit also offers introductory courses in risk management for AI and autonomous systems and houses over 50 fellows involved in international projects, such as adapting audit standards for the European Union and establishing standards for emerging technologies like biometrics.

Training and assessment of auditors include a focus on data inputs and outcomes. Data integrity is apparent in the ForHumanity framework, with attention to the inputs, outcomes, and pipeline. A data pipeline contains the flow of data from its ingestion to a data set used by the AI through its transformation and storage. Auditors use checks and balances to ensure the data meets established metrics and measurements. The hope is that compliance with ForHumanity's' audit will become a "a seal of approval" that will influence buyer behavior over time. To quote founder Ryan Carrier's (2019) philosophy: If we can make good, safe, and responsible AI profitable, whilst making dangerous and irresponsible AIs costly, then we achieve the best possible result for humanity.

This philosophy is rooted in deterrence theory and is modeled on programs like the U.S. Department of Energy's (DOE) Energy Star and Intel Corporation's branding "Intel Inside." Since 1992, the voluntary Energy Star program has become an international standard, with more than 40% of Fortune 500 companies purchasing equipment and 45% of American households knowingly purchasing an ENERGY STAR certified product in 2022 (Environmental Protection Agency, 2023). The "Intel Inside" branding campaign started in 1991, intending to assure non-technical electronics buyers that their choice had quality components. It is easy to see similarities that a ForHumanity certification can make AI tools and applications more acceptable to users.

Auditors certified by ForHumanity learn about the relevant legal frameworks that regulate and legislate bias. For example, AI applications that are used in admissions decisions in US colleges and universities can no longer consider race as a data point. In June 2023, the US Supreme Court ruled that Harvard University and the

University of North Carolina violated the 14th Amendment of the US Constitution and Title VI of the Civil Rights Act of 1964. ForHumanity's training and assessments are also designed with European Union regulations.

The NIST AI Framework was introduced in January 2023. It is prescriptive around four pillars. The first pillar is governance, where the organization sets the overall direction and policy for its use of AI. The framework calls for organizations to identify roles, responsibilities, and resources for mitigating risks. The Mapping pillar is where the potential risks are identified, and their likelihood and impacts are quantified. The framework also includes Measure, a pillar in which data about AI risks are monitored and tracked over time. The fourth pillar is Act, which prescribes the management of risks via controls, monitoring, and adjustments. The NIST AI RMF identifies three categories of bias in AI: systemic, statistical, and human. It identifies three broad challenges: datasets, testing and evaluation, and human factors. It also introduces preliminary guidance for addressing bias in AI applications.

In the popular Harry Potter series of books, students at Hogwarts School of Witchcraft and Wizardry are sorted into their "best-fit" houses by a sorting hat that can sing, talk, and look into the students' minds (Rowling, 1997). This can be likened to an AI application assessing students and assigning them to classes that best align with their capabilities. Having the ability to screen and sort candidates to a pathway that best aligns with their needs may positively impact student retention and success rates while maintaining academic standards at a lower administrative overhead cost than traditional methods, which rely on student transcripts, essays, assessment testing, and the evaluation of those artifacts by staff members. Based on results provided by the AI application, institutions may be able to tailor programs to advanced students or to filter students to classes where they may have additional opportunities or alternative pathways to success (Pallathadka et al., 2023).

In the example of using AI for candidate acceptance or placement within programs, depending on how the AI model is trained, human bias may either be removed from the screening process, and the resulting pool could be a more diversified mixture of candidates or learning bias could negatively influence the results. An AI algorithm will yield results that are systemically prejudiced due to inaccurate assumptions in the ML process. This bias can be injected into algorithms unconsciously or consciously by the systems developers and administrators that develop the ML systems or due to flawed data sets used to sequence the systems (Verma, 2023).

Data persistence can cause AI applications to be biased in their recommendations. Data sets that were created before the afore-mentioned United States Supreme Court decision on college admissions may involve rules that include race as an attribute. This brings into question the longevity of an application's predictive power. Data sets will have to be regularly culled and updated.

AI application developers may repurpose data sets and use the same data in training sets for multiple applications to capture economies of scale. This may cause the system to make Incorrect correlations, resulting in bad decisions or recommendations. This same risk is present in the case of data spillover among training sets used in different applications. Taken out of context, the data that is spilled over may result in bad decisions or recommendations.

4. Conclusion

The biggest challenge for the field of AI may be finding the balance between protecting privacy and restraining advancement. Higher education institutions will benefit from allocating costs towards incorporating measures that consider how to properly engage allowable third-party AI applications securely for student, staff, and administrator use. Faculty need to assess the risks from academia/industry partnerships that are prevalent in the AI field. These partnerships may take the form of subscribing to datasets from third-party providers, which goes back to the prior discussion on transparency and data repurposing. While this paper addresses inherent risks, there are also opportunities, such as cooperation on or funding the research and development of AI applications for higher education. As noted, this discussion shows that higher education has a unique relationship with AI. This is due to the dual uses of AI in higher education, academic and administrative.

Developing a Risk Management Framework per published standards may not seem feasible given the staff and budget constraints that many higher education institutions face. However, it is possible to incorporate aspects of an RMF into institutions' technology policies and decision-making. As of the mid-year 2024, 200 information technology specialists held at least one ForHumanity certification. University and college administrators could engage with a consultant to audit the implementation of AI applications.

Another example is the Govern pillar of the NIST AI RMF. This pillar calls for identifying ownership, writing, and enforcing policies for acceptable use. Schaeffer, Dehghanpour, and Olson (2024) analyze university and college policies about using generative AI in the classroom. They found that many policies give ownership to individual faculty members, who may sanction or prohibit the use. Misuse by students was viewed as a violation of academic integrity in most policies. Managing the risks can build on already defined measures, such as the consequences of academic dishonesty.

The risks of utilizing AI applications in higher education settings are profound, especially risks of bias. Therefore, we recommend that administrators decide on the AI applications and tools based on desired outcomes. Typical applications such as AI as a supplement to classroom activities or in decision-making for admissions, financial aid allocation, or placement, or those AI applications used for recruiting and hiring decisions, require considering the risks and benefits. The benefits that AI applications offer higher education will only be accrued when institutions mandate governance and compliance for the AI applications their stakeholders choose to use.

References

Anonymous 007. (2023). Information Security Confidentiality Geeks for Geeks. https://www.geeksforgeeks.org/information-security-confidentiality/

Appel, G., Neelbauer, J. and Schweidel, D.A., 2023. Generative AI has an intellectual property problem. *Harvard Business Review*, 7.

Armilis. (2023). Data integrity in industry 4.0. https://armilis.com/blog/data-integrity-in-industry-4-0/

Barker, E.B. and Kelsey, J.M., 2007. Recommendation for random number generation using deterministic random bit generators (revised) (pp. 800-900). Washington, DC, USA: US Department of Commerce, Technology Administration, National Institute of Standards and Technology, Computer Security Division, Information Technology Laboratory.

Boyens, J., Smith, A., Bartol, N., Winkler, K., Holbrook, A. and Fallon, M., 2021. Cyber Supply Chain Risk Management Practices for Systems and Organizations (No. NIST Special Publication (SP) 800-161 Rev. 1 (Draft)). National Institute of Standards and Technology.

Burt, A., 2019. The AI transparency paradox. *Harvard Business Review*, 13.

Carrier, R., 2019. Implementing guidelines for governance, oversight of AI, and automation. *Communications of the ACM*, 62(5), pp.12-13.

Dempsey, K.L., Johnson, L.A., Scholl, M.A., Stine, K.M., Jones, A.C., Orebaugh, A., Chawla, N.S. and Johnston, R., 2011. Information security continuous monitoring (ISCM) for federal information systems and organizations.

Environmental Protection Agency Office of Air and Radiation, Climate Protection Partnerships Division. (2023). *National Awareness of ENERGY STAR® for 2022: Analysis of 2022 CEE Household Survey*

Naraine, R. 2023. Microsoft AI Researchers Expose 38TB of Data, Including Keys, Passwords and Internal Messages, *Security Week.* https://www.securityweek.com/microsoft-ai-researchers-expose-38tb-of-data-including-keys-passwords-and-internal-messages/

National Science and Technology Council (US). Select Committee on Artificial Intelligence, 2019. The national artificial intelligence research and development strategic plan: 2019 update (p. 0050). National Science and Technology Council (US), Select Committee on Artificial Intelligence.

Nieles, M., Dempsey, K. and Pillitteri, V.Y., 2017. An introduction to information security. NIST special publication, 800(12).

Oldehoeft, A.E., 1992. Foundations of a security policy for use of the National Research and Educational Network. US Department of Commerce, National Institute of Standards and Technology.

Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J. and Phasinam, K., 2023. Classification and prediction of student performance data using various machine learning algorithms. *Materials today: proceedings*, 80, pp.3782-3785.

Pavlou, C. (2023). Availability bias in the AI era: Understanding its impact on decision-making. https://www.talentlms.com/blog/availability-bias-ai-era/

Pearce, G., 2021. Beware the privacy violations in artificial intelligence applications. ISACA Foundation.

Powell, M., Brule, J., Pease, M., Stouffer, K., Tang, C., Zimmerman, T., Deane, C., Hoyt, J., Raguso, M., Sherule, A. and Zheng, K., 2022. Protecting Information and System Integrity in Industrial Control System Environments.

Pub, F.I.P.S., 2006. Minimum security requirements for federal information and information systems.

Scholl, M.A., Stine, K., Hash, J., Bowen, P., Johnson, L.A., Smith, C.D. and Steinberg, D., 2008. An introductory resource guide for implementing the health insurance portability and accountability act (HIPAA) security rule

Rowling, J. K. 1997*. Harry Potter and the Philosopher's Stone.* Bloomsbury Pub Limited.

Schaeffer, D. M., M. Dehghanpour, and P.C. Olson. 2024. Risk management framework for artificial intelligence in the classroom. *Proceedings of the International Association of Computer Information Systems.*

Verma, R. (2023). Mitigating AI Bias: The Key To A Better Future For Businesses And Society. *Forbes.* https://www.forbes.com/sites/forbesbusinesscouncil/2023/09/13/mitigating-ai-bias-the-key-to-a-better-future-for-businesses-and-society/?sh=2e904a895a93

Using ChatGPT in Teaching Computer Programming and Studying its Impact on Students Performance

Shubair Abdulla[1], Sameh Ismail[1,2], Yasser Fawzy[1,3] and Abdelrahman Elhaj[1]
[1]Sultan Qaboos University, Oman
[2]Cairo University, Egypt
[3]Helwan University, Egypt

shubair@squ.edu.om
samehsaid@squ.edu.om
yasser.fawzy@squ.edu.om
abdoelhaj@squ.edu.om

https://doi.org/10.34190/ejel.22.6.3380

Abstract: Given the recent emergence of artificial intelligence as an important topic that can contribute to improve curricula and lecture delivery, an increasing number of scholars are investigating its impact in various educational fields. The ChatGPT, which is an artificial intelligence model developed by OpenAI, represents a significant advancement in the generative artificial intelligence area. Since its announcement, integrating the ChatGPT into computer programming curricula - and into other scientific curricula - has yielded some challenges. The main challenge is clearly evident in the conduct of in-class tests and in-lab assignments, where the students are given specific tasks to be accomplished within a certain time frame. As they might seek help from ChatGPT, these types of assessments could be considered a potential threat to academic integrity and may be viewed as a form of academic dishonesty. This study aims at integrating ChatGPT into computer programming curricula, exploring its potential to enhance undergraduate education. It follows a mixed methods approach to examine the potential integration of ChatGPT in teaching computer programming as a supplementary tool. A quasi-experimental design is followed, in which an experimental group is allowed to use ChatGPT and compared to a control group that was not. A research sample of 26 undergraduate students (13 males, 13 females) from the College of Education at Sultan Qaboos University participated in the study. The methodology encompassed three research instruments: in-class exams, in-lab assignments, and semi-structured interviews. These research instruments were utilized to assess the impact of ChatGPT on students' academic performance, which served as both independent (use of ChatGPT) and dependent (student performance) variables. The quantitative analysis revealed a significant enhancement in students' performance, while the qualitative analysis of semi-structured interviews indicated that participants view ChatGPT as a valuable support for learning. Feedback from participants suggested combining ChatGPT with traditional teaching methods to optimize learning outcomes. This study highlights the feasibility and educational benefits of incorporating AI tools like ChatGPT into teaching methodologies. It suggests that such integration can provide a more engaging and effective learning environment, potentially revolutionizing computer programming education. This paper supports e-learning practice by integrating AI-driven tools like ChatGPT into the educational framework and advances the e-learning area by demonstrating these technologies' potential to improve student academic performance in the learning environments. However, the study also acknowledges the need for further research to explore the long-term effects of AI integration in educational settings and to address any emerging challenges. These findings propose a promising direction for future curricular enhancements and suggest an effective method for the integration of AI technologies to support and enrich traditional educational frameworks.

Keywords: ChatGPT, Teaching computer programming, In-class exams, In-lab assignments, Mixed-methods approach

1. Introduction

Computers, smart devices and robots have entered almost all aspects of human life. These devices are currently being utilized in communication, commerce, marketing, teaching and learning, in addition to many other aspects. The importance of programming is evident in building communication interfaces, controlling these digital devices, and managing their systems. The demand for programmers has increased significantly (National Academies of Sciences, Engineering, & Medicine, 2018), and teaching programming has become one of the main goals (Kanika, Chakraverty, & Chakraborty, 2020) of higher education institutions to prepare graduates with programming skills to meet this growing demand. The endeavor of higher education institutions to equip graduates with distinguished programming knowledge and skills has been reflected in the development of curricula and methods for teaching computer programming (Erümit, 2020). As a result, significant efforts have been dedicated to developing all aspects of the educational process. Among the aspects involved in the development are methods for assessing students. In teaching computer programming, assessment plays a crucial role. It enables teachers to evaluate student’s abilities to recall facts and basic terms, to measure their

Cite this article: Abdulla, S et al. 2024. “Using ChatGPT in Teaching Computer Programming and Studying its Impact on Students Performance”, *Electronic Journal of e-Learning*, 22(6), pp 66-81, https://doi.org/10.34190/ejel.22.6.3380

understanding of codes, and to assess their ability to solve problems using the knowledge and understanding (Erümit, 2020). Moreover, assessing computer programming contributes to developing students' programming skills and enhancing their higher-order thinking (Tikva & Tambouris, 2021).

Artificial Intelligence (AI) has recently become an essential part of education. Many applications of AI have emerged and had a positive impact on teaching, learning, curricula, and assessment (Goksel & Bozkurt, 2019). Among of these is ChatGPT, that is an AI-powered application developed by OpenAI Company. The ChatGPT provides a wide range of services that could be utilized by students such as study materials and in-lab assignments help (Haleem, Javaid & Singh, 2022). However, since its announcement, integrating the ChatGPT into computer programming curricula - and into other scientific curricula - has become a topic of debate (Ali et al., 2023). Some educators consider it a potential threat to academic integrity and others argue that its use in computer programming curricula may be viewed as a form of academic dishonesty (Zielinski et al., 2023). This problem is clearly evident in the conduct of some assessment methods such as in-class tests and in-lab assignments, where the students are given specific tasks to be accomplished within a certain time frame. Examples of tasks as in-class tests and in-lab assignments are analyzing and evaluating existing software, writing programs to solve real-world problems, or researching programming languages and creating reports or presentations (Şenel & Şenel, 2021).

The main goal of this study is to examine the integration of ChatGPT into computer programming curricula. The study attempts to measure the impact of using ChatGPT on students' academic performance and its findings assist educators in making decisions about integrating AI online platforms into curricula. The research questions that this study attempt to address are as follows:

- How does the use of ChatGPT affect students' performance in learning computer programming?
- What are the students' perceptions regarding the use of ChatGPT in learning computer programming?
- What is the effective method for integrating ChatGPT into a computer programming curricula?

To explore the impact of ChatGPT in computer programming education, a quasi-experimental design and semi-structured interviews were incorporated in this study. We compared over 8 weeks the performance of experimental group (allowed to use ChatGPT) with a control group (not allowed to use ChatGPT) utilizing two instruments, in-class exams and in-lab assignments. Participants included 26 second-year undergraduate students of Instructional and Learning Technology major in Sultan Qaboos University (SQU). In order to provide a comprehensive assessment of ChatGPT's effectiveness, a specialized programming course that utilized C# programming language in the MS Visual Studio environment was utilized in this study.

To ensure the integrity of the research and verify that the students in the control group did not use ChatGPT, several measures were implemented. First, students were informed about the research process and the importance of adhering to the research integrity recommendations. We also emphasized that their commitment to not using ChatGPT was essential for producing real and reliable results. Finally, proctoring was conducted during all in-class exams and in-lab assignments to monitor students' activities in the control group.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature and theoretical frameworks related to the integration of ChatGPT into the computer programming curricula. Section 3 describes the methodology used to collect and analyze data in order to address the research question. Section 4 presents the results of the study along with a discussion of their implications. Finally, Section 5 concludes the study by presenting the key findings, implications, and recommendations for future research.

2. Literature Review

2.1 Using ChatGPT in Teaching Computer Programming

The accuracy of ChatGPT in responding the questions in various fields of science has been widely recognized and led to increase its use in teaching and learning. Its ability to understand and respond to context makes it a valuable learning tool for students (Hassani & Silva, 2023). However, the use of ChatGPT has raised debate about its potential effects on traditional teaching methods. The main issue is that the use of ChatGPT technology may pose a potential threat to academic integrity. Some studies have found that students have begun using ChatGPT and relying more heavily on it to complete essay assignments (Rosenberg, 2023).

Several research studies have examined the use of ChatGPT in diverse fields of education. For instance, Rudolph, Tan, and Tan (2023) discussed the future prospects of ChatGPT in higher education. Their study analyzed its applications for students and teachers, and summarized the potential benefits and risks involved. Similarly, Baidoo-Anu & Owusu Ansah, L. (2023) recommended the collaborative efforts of policymakers, educators, researchers, and technology experts to make the most of ChatGPT's potential for enhancing education and supporting student learning. Additionally, Yan (2023) explored the implications of ChatGPT on students' writing behavior and learning. His study recommended developing regulatory policies and redefinition of plagiarism to ensure responsible use of ChatGPT.

There are a few papers published to explore the integration of ChatGPT into computer programming curricula. For example, Qureshi (2023) addressed in his paper the use of ChatGPT as a mean of teaching and learning fundamental programming courses in undergraduate Computer Science curricula. The paper discussed the benefits and obstacles of using ChatGPT in this context and examined its impact on students' learning outcomes. The paper findings suggested that students using ChatGPT had an advantage, but there were inconsistencies and inaccuracies in their submitted code, which negatively affected their overall performance and presents challenges that must be proactively managed with an ethical stance. Another paper published by Joshi et al., (2023) aimed to explore the potential impact of ChatGPT on academic integrity in undergraduate computer science education by investigate whether students may be tempted to use ChatGPT to complete take-home assignments and exams, and whether this can negatively affect their learning outcomes. The paper concluded by providing constructive recommendations to both students and instructors on how to use ChatGPT ethically and effectively in academic settings. The authors recommended that instructors educate their students on the potential drawbacks and limitations of ChatGPT and encourage them to use it as a supplementary tool for learning rather than a substitute for acquiring knowledge.

Up to date, the integration of ChatGPT into computer programming curricula has not received much attention in contrast with other educational curricula. Therefore, there is a need for more research to explore the feasibility, effectiveness, and impact of integrating ChatGPT into computer programming curricula, as well as to identify the best practices and strategies for leveraging this technology to enhance students' learning outcomes.

In addition to direct coding assistance, the Generative AI in education provides other tools designed to enrich learning experiences across multiple aspects of computer science education. For instance, Wolfram Alpha (Abramovich, 2021) offers computational intelligence that can serve as a valuable tool for understanding the mathematical foundations of computer science. Another example, Google's BERT, a natural language processing (NLP) mode, can facilitate the development of interactive educational chatbots, offering personalized tutoring and support in learning programming concepts. Moreover, there are several other Generative AI environments that could be utilized for enhancing teaching and learning. These include tools like Codex, developed by OpenAI, which offers capabilities to assist in code generation and debugging. Another notable environment is DeepMind's AlphaCode (Lertbanjongngam et al., 2022) designed to tackle competitive programming challenges and promote problem-solving skills and understanding of algorithms among students. Integrating these Gen AI tools into computer programming curricula can offer diverse perspectives and methodologies for learning and enriching the students' learning experience.

2.2 In-Class Exams and In-Lab Assignments

The in-class exams and in-lab assignments are commonly used assessment tools. They refer to exams and assignments given to students within regular class time. Like scientific curricula, these types of assessment are used in computer programming curricula to evaluate students' understanding of the concepts as well as their ability to apply what they have learned (Bengtsson, 2019). The components of each of these two types of assessments vary; for example, an in-class examination may include written tests and assignments, while an in-lab assignment usually involves a specific task that must be completed within the computer lab. However, these two types are similar in that each is designed to assess a student's mastery of course material (Spiegel & Nivette, 2023).

The use of in-class exams and in-lab assignments along with their impact on the students' academic performance have been addressed in previous studies. For example, Erhel & Jamet (2013) concluded in their research that a regular and timely feedback on in-lab assignments can improve students' understanding of course concepts and increase their motivation to learn. On the other hand, some research shown that the in-class exams can provide a more authentic assessment of students' abilities and allow students to work at their own pace and in a more relaxed environment (Bengtsson, 2019). However, there are some challenges associated with the use of in-class exams and in-lab assignments. Regarding the in-lab assignment, Wang et al. (2012) stated that some students struggle to manage their time effectively, and this often results in last-minute submissions with incorrect answers. The unequal opportunities students to learn is another challenge. For different reasons, some students may not have access to the necessary resources or technology to complete in-class exams and in-lab assignments

(Shakeel et al., 2021). The emergence of ChatGPT has caused another challenge. Some research has pointed to ChatGPT as a potential threat to academic integrity and stated that its use by students for assignments or exams is a form of academic dishonesty (Cotton, Cotton & Shipway, 2023). Despite the challenges associated with their use, the in-class exams and in-lab assignments have a significant impact on students' academic performance. However, further research is needed to explore best practices for using ChatGPT in teaching and learning and to determine the effective ways to support students.

This study targets computer programming students focusing on their performance and perceptions on the integration of ChatGPT into computer programming curricula. Additionally, it attempts to find an appropriate way to integrate ChatGPT as a support learning tool. There are two reasons for this focus. First, ChatGPT has been increasingly used in different fields of education. However, there are still concerns about academic integrity and the appropriate methods to deal with this emerging technology (Joshi et al. 2023). Second, most of the previous studies that examined the integration of ChatGPT into the education field are observational studies and do not involve experiments. Therefore, more research should be conducted to provide more empirical findings (Sallam, 2023).

3. Method

3.1 Research Design

This study utilized a mixed methods approach to investigate the effects of implementing ChatGPT in teaching computer programming on student outcomes. Specifically, a quasi-experimental design was adopted, wherein an experimental group exposed to ChatGPT was juxtaposed against a control group that was not. The study was conducted over a period of 8 weeks, employing three distinct research instruments: in-class examinations, in-lab assignments, and semi-structured interviews. The independent variable in this study was the use of ChatGPT in teaching computer programming. Conversely, the dependent variables encompassed student performance, as evaluated through in-class exams and in-lab tasks.

In light of our experimental design, the following null hypotheses will be tested to find out the impact of ChatGPT on learning outcomes in computer programming education:

- $H_{0:}$ There is no significant difference in the comprehensive programming concept between the experimental group and the control group.
- $H_{1:}$ There is no significant difference in the ability to apply programming concepts in hands-on scenarios between the experimental group and the control group.

3.2 Theoretical Framework

This study aims to examine the integration of ChatGPT into computer programming curricula and assess its impact on students' academic performance. To achieve its objectives, this study applies the Unified Theory of Acceptance and Use of Technology (UTAUT) model as a theoretical framework (Venkatesh et al., 2003). The UTAUT framework was developed to examine the use of information systems depending on four key constructs: (1) Performance Expectancy, (2) Effort Expectancy, (3) Social Influence, and (4) Facilitating Conditions. The performance expectancy factor refers to the impact of technology use on improving student's performance. In the context of using ChatGPT in teaching computer programming, this could include the improvement problem-solving abilities, and understanding of programming concepts. The effort expectancy factor refers to level of convenience and ease of technology use. This might involve the ease of interacting with ChatGPT and the user-friendliness of the interface in the context of using ChatGPT in teaching computer programming. The social influence factor measures the levels of importance of technology. In the case of using ChatGPT in teaching computer programming, this might involve the expectations or recommendations from students. Finally, the facilitating conditions factor refers to the availability of an organizational infrastructure to support the use of the technology. With regards of using ChatGPT, this might include the availability of procedures to explain how to use ChatGPT in teaching computer programming.

3.3 Participants

This study was conducted by a quasi-experimental design. A cohort of 26 second-year undergraduate students participated, comprising an equal gender distribution with 13 males and 13 females. All participants were enrolled in the Instructional and Learning Technology major and were undertaking a one-semester course, "Instructional Computer Programming (I)," at the College of Education, Sultan Qaboos University. The research outcomes may be influenced by the academic background of participants. Their familiarity with instructional technology could make them to more effectively engage with ChatGPT, which might result a more positive reception. The course structure involved two face-to-face classes each week, with each session lasting two hours. These classes were conducted in computer-equipped classrooms. The curricula covered topics related to visual computer programming, including but not limited to: variables, decision statements, iteration statements, classes and objects, arrays, addressing syntax and logic errors, and reading and writing files. Evaluation methods encompassed paper exams, in-class exams, in-lab tasks, and homework assignments. To ensure ethical considerations, students were informed about the study's objectives prior to its commencement, and their informed consent was sought. All students of the course were exposed to equitable learning experiences as all of them consented to participate, which means that our sample is a convenient sample. The sample size of 26 participants represents the total number of enrolled students in the course, which represents the maximum feasible sample for this study. Since, all registered students participated, extending the sample size beyond this number was not possible.

3.4 Procedures

The Instructional Computer Programming (I) course primarily utilized visual programming language within the MS Visual Studio environment. It provides students with a comprehensive platform for learning programming concepts and applying them using C# programming language. This setup ensured a standardized technological context for evaluating the impact of ChatGPT assistance. Data collection covered eight weeks of the SP2023 semester, initiating in the 2nd week and concluding in the 9th week. Quantitative data collection was focused on weeks 2 to 7, while qualitative data were gleaned during weeks 8 and 9. With regards to quantitative data collection, two instruments were employed: in-class exams and in-lab assignments. Each instrument had two distinct versions: for example, in-class exam (1) and in-class exam (2) for the in-class assessment; in-lab assignment (1) and in-lab assignment (2) for the in-lab assignment. The in-class exams assessed students' comprehension of programming concepts, whereas the in-lab assignments tested their capability to apply these concepts in hands-on programming scenarios. Each instrument was deployed twice. The initial deployment split students into two subsets: an experimental group (with access to ChatGPT) and a control group (without ChatGPT access). The same split was followed during the deployment of in-class exam (1). However, in the subsequent in-class exam (2), roles were swapped: students in the original experimental group moved to the control group and vice versa. The same system was applied in the administration of the in-lab assignments, ensuring that each participant experienced both conditions. In addition to offering both groups equitable exposure to ChatGPT, this procedure also minimized potential confounding variables and met ethical guidelines.

Regarding qualitative data, semi-structured interviews were conducted with 16 students selected from the total sample of 26. These students were chosen based on their levels of engagement with ChatGPT and their willingness to provide in-depth feedback, as assessed during in-lab assignments 1 and 2. These interviews aimed to clarify students' perceptions about the efficacy, convenience, and user-friendliness of leveraging ChatGPT in solving in-class exams and in-lab assignments. Although a set of predefined questions guided the interviews, the interviewer allowed flexibility for follow-up questions, elaboration, and exploration of new topics based on the participants' responses.

3.5 Instruments

Three research instruments were used in this study to collect both quantitative and qualitative data:

1. In-class exams: These exams were administered twice: first in the 4th week and later in the 7th week. The initial exam aimed to assess participants' comprehension of the concepts related to graphical user interface, event handling, and variables and constants, while the second focused on functions and classes, and sub-procedures. Both tests presented essay-style questions, demanding written answers to demonstrate their understanding of the topic. The questions were developed based on the learning objectives and content of Instructional Computer Programming (I) course. Figures 1 and 2 show the in-class exams used as instruments research in this study.
2. In-lab assignments: These tasks were assigned on two occasions: in the 3rd week and the 6th week. The first assignment tasked participants with a decision-making exercise, specifically focusing on the 'If-blocks' concept. The subsequent assignment directed participants to solve problems related to arrays and file handling. Participants were instructed to design and develop complete programs and submit their code solutions for evaluation. The tasks were designed to align with the learning objectives and content. Figures 3 and 4 present the in-lab assignment used in this research.
3. Semi-structured interviews: Semi-structured interviews were conducted with a select group of participants to collect their perceptions of using ChatGPT, specifically its convenience and efficacy in

approaching in-class exams and in-lab assignments. The interview framework was established based on UTAUT framework and a thorough literature review. A set of 5 questions (shown in Table 1) was used to guide the interviews. Prior to the main study, a pilot test was conducted with a smaller participant group of 4 students to fine-tune the interview questions.

TECH3222 - Instructional Computer Programming I

In-class exams (1)

An application has been created to convert one kilometer to inches, feet, yards, and miles as in the screenshot below. When a button is clicked, the label has an equivalent number to one kilo, and when the 'Exit' button is clicked, the application terminates. You are required to write the click event for all five buttons.

Tip:

1 Kilometer = 3,281 feet, 1 Kilometer = 39,370 inches, 1 Kilometer = 0.6214 miles, and 1 Kilometer = 1,093.6 yards.

Assessment Rubric

#	Criteria	Unsatisfactory (0)	Needs Improvement (1)	Good (2)
1	Inches button click event			
2	Feet button click event			
3	Yards button click event			
4	Miles button click event			
5	Exit button click event			
			Total Earned Points:	

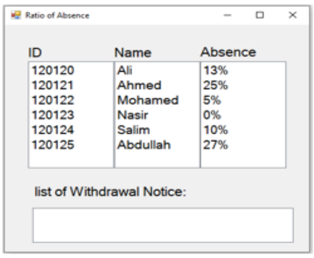
Figure 1: The first in-class exam - variables and constants concepts

TECH3222 - Instructional Computer Programming I

In-class exams (2)

Design a program that checks the ratio of absence for all students and shows on the form a list "Absentee Withdrawal Notice" for all the students with ratio more than 20%. The design and input are shown below

Ratio of Absence for the Students



ID:	120120	120121	120122	120123	120124	120125
Name:	Ali	Ahmed	Mohamed	Nasir	Salim	Abdullah
absence:	13%	25%	5%	0%	10%	27%

Assessment Rubric

#	Criteria	Unsatisfactory (0)	Needs Improvement (1)	Good (2)
1	Input and output			
2	Sub procedures			
3	functions			
4	Loop			
5	Conditions			
			Total Earned Points:	

Figure 2: The second in-class exam - functions and sub procedures concepts

TECH3222 - Instructional Computer Programming I

In-Lab Assignments (1)

Telling which part of the day it is

You are given the task of designing a program that tells you which part of the day it is, given a user-entered time (hour & minute). Please use the 24h clock. Show the user two input text boxes, one for the hour of day, the other for the minute. The user then clicks a "Convert" button, your program should display the corresponding time of day in the color specified below. Use the following table:

Hour	Time of Day	Color
21:00-23:59 or 0:00-5:59	Night	Black
6:00-10:59	Morning	Yellow
11:00-13:59	Lunch	Orange
14:00-17:59	Afternoon	Light Blue
18:00-20:59	Evening	Dark Blue

Assessment Rubric

#	Criteria	Unsatisfactory 0	Needs Improvement 1	Good 2
1	Graphical User Input (GUI) Design			
2	Using Datatype Conversion			
3	Using If Statements and select case block			
4	Validation of User Inputs			
5	The comparison between if then elseif block and select case block			
			Total Earned Points:	

Figure 3: The first in-lab assignment - decision (If-blocks) concepts

TECH3222 - Instructional Computer Programming I

In-Lab Assignments (2)

Students' Grade book

Develop a VB.NET program that reads from a CSV file (st_performance.csv) a list of student marks. The list contains 28 students of programming course section 10, and stores the midterm mark (20%), assignment 1, 2, 3, and 4 marks (10% each), and final exam mark (40%). An example is shown in the image below.

```
Ahmed,16.5,10,8,6,10,33.5
Amjad,19.5,10,10,10,9,36
Basim,15,7.5,7.5,8,9,37
Hashim,12,8,7,6,9,30
Qusay,15.5,10,10,10,9,36.25
```

The program then defines sub programs or functions for:

1. Calculates and displays the total score for each student
2. Calculates and displays the grade for each student (A, A-, B+, B, …)
3. Calculate and display the percentage of success and failure in the section
4. Calculate and display the percentage of success and failure in the final exam
5. Calculate and display the percentage of grades A, A-, B+, B, …
6. Calculate and display the maximum mark
7. Calculate and display the minimum mark

The program must provide an option to save the calculation results to a text file. The image below shows an example of saving the results of calculations.

Assessment Rubric

#	Criteria	Unsatisfactory 0	Needs Improvement 1 or 0.5	Good 2 or 1
1	Calculates the total score for each student **(1 marks)**			
2	Calculates the grade for each student (A, A-, B+, B, …) **(2 marks)**			
3	Calculate the percentage of success and failure in the section **(2 marks)**			
4	Calculate the percentage of success and failure in the final exam **(2 marks)**			
5	Calculate the percentage of grades A, A-, B+, B, … **(1 marks)**			
6	Calculate the maximum mark **(1 marks)**			
7	Calculate the minimum mark **(1 marks)**			
			Total Earned Points:	

Figure 4: The second in-lab assignment - arrays and files concepts

Table 1: Question of the Semi-structured interview

#	Question
1	What types of tasks and challenges do you typically try to overcome when utilizing ChatGPT for your learning in computer programming?
2	In what ways has ChatGPT been helpful in your learning journey of computer programming?
3	Have you encountered any limitations or drawbacks when using ChatGPT for learning programming?
4	How do you think ChatGPT compares to more traditional methods of learning programming, such as textbooks, online tutorials, or instructor-led classes?
5	Do you believe ChatGPT should be integrated into computer programming education?

3.6 Validity and Reliability

To ensure a high degree of research dependability, several measures were implemented to reinforce validity and reliability. For the in-class exams and in-lab assignments, content validity was assessed. All versions of these instruments were initially designed by the course instructor. The instruments were reviewed by two course coordinators to confirm the appropriateness of their content. An additional measure of reliability, the test-retest method, was applied to the in-class exams and in-lab assignments. The goal of this method was to assess the consistency of these instruments. Since these instruments are parts of student grades, the test-retest was conducted on 10 students from different cohorts that are separate from the main study group. A two-week interval was set between the initial test and its subsequent retest. After recording the scores from both tests, a correlation coefficient was calculated, yielding a value of 0.73. This result suggests that the in-class exams and in-lab assignments exhibit an acceptable level of reliability. As for the semi-structured interview instrument, two specialists in the domain reviewed the instrument, offering their perspectives and recommendations. Based on their feedback, some modifications were made to the interview questions, which included amending three questions and deleting one question.

3.7 Data Analysis

The quantitative data analysis in this study involved the examination of student's performance from the in-class exam and the in-lab-assignment instruments. The purpose of data analysis was to assess the impact of using ChatGPT on students' performance in computer programming. Specifically, the data analysis process focused on comparing the performance of two groups: the experimental group (using ChatGPT) and the control group (not using ChatGPT). The raw data collected from the in-class exams and in-lab assignments was gathered and organized for analysis. This data consists of the scores obtained by students. The students' scores in these two exams are set at 100 to align with the evaluation system used in the Instructional Computer Programming (I) course. The data was thoroughly examined to ensure completeness, accuracy, and consistency.

Descriptive statistics were calculated for the in-class exam scores and in-lab assignment scores separately for both the experimental and control groups. Measures such as mean, standard deviation, and range were computed for each group to provide an overview of their performance. A comparative analysis was conducted to assess the differences in performance between the experimental and control groups. Initially, the Shapiro-Wilk test was performed for each group to assess normality. Based on the results of the Shapiro-Wilk tests, statistical tests such as independent samples t-tests or Mann-Whitney U tests (Royston, 1992; Sawilowsky, 2005) were conducted to determine if there were significant differences in the exam scores and assignment scores between the two groups.

Descriptive statistics were generated using MS Excel, while the Python programming language facilitated the comparative analysis. Specifically, within Python, the Shapiro-Wilk test, independent samples t-test, Mann-Whitney U test, and calculation of Cohen's d effect size measure were conducted. The qualitative data were collected through semi-structured interviews conducted with a subset of participants. The interviews aimed to explore the students' perceptions regarding the convenience and ease of using ChatGPT for solving in-class exams and in-lab assignments. The data analysis process for the qualitative data involved the following steps:

- The transcriptions were reviewed to develop a deeper understanding of the participants' perspectives.
- A thematic coding approach (Shoufan, A. 2023) was employed to identify patterns, themes, and concepts within the data.

- A coding scheme was developed based on the identified themes and patterns. The coding scheme provided a structured framework for organizing and categorizing the data.
- The coded data were summarized, condensed, and organized to capture the essential information related to the research questions. This process involved the extraction of relevant quotes, illustrative examples, or excerpts that represented the identified themes and supported the interpretation of the findings.

The qualitative data analysis process was facilitated by the use of MS Word, a word processor to aid in organizing, coding, and managing the data. The program allowed the efficient arrangement of the text and encrypted passages in tables in order to carry out the comprehensive analysis process.

4. Results and Discussion

The primary objective of this research is to examine the integration of ChatGPT into computer programming curricula and assess its impact on students' academic performance. Guiding this investigation are the following research questions:

- How does the use of ChatGPT affect students' performance in learning computer programming?
- What are the students' perceptions regarding the use of ChatGPT in learning computer programming?
- What is the effective method for integrating ChatGPT into a computer programming curricula?

The following null hypotheses were also formulated:

- H_0: There is no significant difference in the comprehensive programming concept between the experimental group and the control group.
- H_1: There is no significant difference in the ability to apply programming concepts in hands-on scenarios between the experimental group and the control group.

The subsequent sections will present and discuss findings pertinent to these questions.

4.1 Effect of ChatGPT on the Performance of Computer Programming Leaners

4.1.1 Descriptive statistical analysis

To provide a snapshot of the overall performance, Table 2 presents the mean, standard deviation, minimum, and maximum scores for both groups across the two tests, the in-class exam and the in-lab-assignment. In the first in-class exam, the experimental group averaged 78.85 compared to the control's 58.08. The smallest difference between means is seen in the second in-class exam, with scores of 62.31 for the experimental group and 63.85 for the control group. Additionally, the range of scores indicates a generally wider variation in the control group, especially in the second in-class exam, where the maximum score reached 90 compared to the experimental group's 80. Table 3 shows details about the individual scores for application 1 and application 2 for the In-Class Exam and In-Lab Assignment. Scores within the control (C) and experimental (E) groups varied. For Application#1 exams, control group scores mostly ranged between 40 and 70, while the experimental group often surpassed 90. The experimental group generally scored higher than the control in in-lab assignments, particularly for Application#2.

Table 2: Descriptive statistics of scores for both groups across in-class exams and in-lab assignments

Group	Instrument/application	Mean	STD	Range	
				Min	Max
Experimental Group	In-Class Exam (1)	78.85	12.75	57	99
	In-Class Exam (2)	62.31	11.20	40	80
Control Group	In-Class Exam (1)	58.08	12.02	40	80
	In-Class Exam (2)	63.85	15.95	40	90
Experimental Group	In-Lab Assignment (1)	80.77	9.17	70	100
	In-Lab Assignment (2)	84.62	6.34	70	90
Control Group	In-Lab Assignment (1)	69.23	7.30	60	80
	In-Lab Assignment (2)	76.92	7.22	70	90

4.1.2 *Inferential statistics analysis*

While descriptive statistics offer a preliminary insight into the performance of both the experimental and control groups, they do not provide a comprehensive understanding of the underlying differences between the groups. Therefore, inferential statistical analysis is crucial. This section will detail the results of the inferential statistical tests that were conducted to more robustly assess the impact of integrating ChatGPT into computer programming learning. Assessing the normality of score distributions for both groups is a foundational step in inferential statistical analysis. The Shapiro-Wilk test, commonly use to check for normal distribution, suggests the that the data follows a normal distribution. A p-value below a predetermined alpha level (usually set at 0.05) indicates that the data is not normally distributed. Table 4 presents the results of this Shapiro-Wilk test for both the experimental and control groups.

Table 3: Individual scores for application 1 and application 2 in in-class exams and in-lab assignments

Application#1				Application#2			
Student #	**Group**	**In-Class Exam (1)**	**In-Lab Assignment (1)**	**Student #**	**Group**	**In-Class Exam (2)**	**In-Lab Assignment (2)**
1	C	70	60	26	C	70	80
2	C	65	70	25	C	60	80
3	C	50	80	24	C	50	70
4	C	55	60	23	C	80	70
5	C	80	70	22	C	70	90
6	C	70	70	21	C	40	80
7	C	60	70	20	C	90	70
8	C	65	80	19	C	60	70
9	C	50	60	18	C	50	70
10	C	45	70	17	C	40	80
11	C	40	60	16	C	60	80
12	C	65	70	15	C	70	90
13	C	40	80	14	C	90	70
14	E	90	80	13	E	70	80
15	E	58	100	12	E	40	90
16	E	77	80	11	E	60	80
17	E	60	70	10	E	50	90
18	E	99	80	9	E	50	90
19	E	94	80	8	E	70	90
20	E	57	80	7	E	60	90
21	E	80	90	6	E	80	80
22	E	85	70	5	E	60	80
23	E	85	70	4	E	60	90
24	E	80	90	3	E	60	90
25	E	80	70	2	E	70	80
26	E	80	90	1	E	80	70

Table 4: Shapiro-Wilk test results

Tool/Application	P-Values	
	Experimental	Control
In-Class Exam (1)	0.140	0.570
In-Class Exam (2)	0.390	0.430
In-Lab Assignment (1)	0.070	0.010
In-Lab Assignment (2)	0.002	0.005

For the In-Class Exam (1 & 2) and the In-Lab Assignment (1), p-values are above the typical threshold (e.g., 0.05), indicating that the data sets are normally distributed. On the other hand, the p-values for the In-Lab Assignment (2) are below this threshold for both groups, which means that this particular data set is not normally distributed.

Based on the results of the Shapiro-Wilk normality test and in accordance with standard statistical practices, the Independent Samples t-test was chosen for analyzing the scores of In-Class Exam (1 & 2) and In-Lab Assignment (1). Given that the data for In-Lab Assignment (2) did not exhibit normal distribution, the Mann-Whitney U test, a non-parametric alternative to the Independent Samples t-test, was employed for this specific dataset. The results of the inferential statistics are presented in the following:

In-Class Exam (1):

- Independent Samples t-test:
- *t-value: -4.1074*
- *Degrees of Freedom (df): 24*
- *p-value: 0.0004*
- Interpretation: The resulting p-value of 0.0004 is much less than the typical significance threshold of 0.05. The results indicate a statistically significant difference in the scores between the experimental and control groups. The negative t-value suggests that the mean of the experimental group (78.85) is greater than the mean of the control group (58.08) as shown in Table 1. This difference is statistically significant ($p < 0.05$), meaning the experimental group performed significantly better than the control group in the In-Class Exam (1).

In-Class Exam (2):

- Independent Samples t-test:
- *t-value: 0.2734*
- *Degrees of Freedom (df): 24*
- *p-value: 0.7869*
- Interpretation: The resulting p-value of 0.7869 is higher than the typical significance threshold of 0.05. This means there's no statistically significant difference in the scores of the experimental group compared to the control group for this exam. The performance of both groups is statistically similar for In-Class Exam (2).

In-Lab Assignment (1):

- Independent Samples t-test:
- *t-value: -3.4114*
- *Degrees of Freedom (df): 24*
- *p-value: 0.0023*
- Interpretation: The resulting p-value is 0.0023, which is below the typical significance threshold of 0.05. This result indicates a statistically significant difference between the scores of the two groups. The negative t-value indicates that the mean score of the experimental group (80.77) is greater than the mean score of the control group (69.23) as shown in Table 1. Given that the p-value is less than 0.05, this difference is statistically significant. Hence, the experimental group outperformed the control group in In-Lab Assignment (1).

In-Lab Assignment (2):

- Mann-Whitney U test:
- *U-value: 39.5*

- *p-value: 0.0152*
- Interpretation: The U-value here indicates a difference in the ranking of scores between the two groups. Since the mean of the experimental group (84.62) is greater than the mean of the control group (76.92) for the In-Lab Assignment (2) as shown in Table 1 and the p-value is less than 0.05, we can conclude that the experimental group performed significantly better than the control group in In-Lab Assignment (2).

The null hypotheses (H_0 and H_1) suggested that there are no significant differences in performance between the experimental and control groups in terms of in-class exam and in-lab assignment outcomes. The statistical analyses conducted, the Sample t-tests and Mann-Whitney U test led to the rejection of the null hypothesis for the first in-class exam (H_0) and the in-lab assignments (H_1). On the other hand, the null hypothesis was not rejected for the second in-class exam, which suggests that no significant difference in performance between the two groups in that context only.

4.1.3 Discussion

As the results of the descriptive and inferential statistics analyses suggest, the use of the ChatGPT appears to have a positive impact on the performance of computer programming students in the contexts studied. In the In-Class Exam (1), students using ChatGPT significantly outperformed those who did not. This is likely because the content of this exam was more general in nature, and such broad topics (converting kilometers to feet or inches) are areas that ChatGPT is undoubtedly trained on. While the control group primarily depended on their memory and understanding of the concepts, the experimental group depended on ChatGPT for answers which could easily address such general questions. On the other hand, In-Class Exam (2) presented a unique scenario. Despite its focus on functions and sub-procedures, the questions were customized to a specific topic that is: addressing absenteeism at Sultan Qaboos University. Such specialized contexts might not be readily available in ChatGPT's training data, and it would require specific prompting or expertise to get the most out of the ChatGPT. Hence, there was no statistically significant difference between the two groups in this exam. This shows that while ChatGPT can be a valuable learning tool for general programming queries, its effectiveness may vary when faced with highly specific or localized questions. This implies that in certain contexts or situations, the usage of ChatGPT might not necessarily provide a distinct advantage.

In regard to the practical applications observed in In-Lab Assignment (1) and In-Lab Assignment (2), the data indicates that students who used ChatGPT had a clear performance advantage over those who did not. During In-Lab Assignment (1), students were asked to solve a general problem commonly faced by many: determining which part of the day it is using decision-based (If-blocks) concepts. Given the broad and common nature of this assignment, it's reasonable that ChatGPT had been trained on similar queries. The results in In-Lab Assignment (1) is similar to the results observed in In-Class Exam (1). However, In-Lab Assignment (2) introduced a more specific challenge, where students were asked to compute the grade book of students at Sultan Qaboos University. Though the experimental group did outperform the control group, the difference in means (84.62 for the experimental group vs. 76.92 for the control group) is relatively small. This narrow performance gap can be potentially attributed to the fact that students had access to Visual Studio IDE, enabling them to write, debug, and optimize their programs before submission. Such resources may have lessened the distinct advantage offered by ChatGPT in this particular scenario.

4.1.4 Response to the research question: How does the use of ChatGPT affect students' performance in learning computer programming?

In response to this question, based on the results of the descriptive analysis, we can conclude that the influence of the ChatGPT on computer programming students' performance is context dependent. For examinations, if questions entail solving highly customized problems, ChatGPT may not offer a noticeable advantage. In practical tests, while ChatGPT does contribute to enhanced performance, the advantage is slightly reduced since students utilize editors and compilers like Visual Studio IDE to refine their programs before submission. While ChatGPT can be beneficial for general programming queries, its impact decreases in situations demanding specialized solutions or when other optimization tools are accessible to students.

4.2 Students' Perceptions Regarding the Use of ChatGPT in Learning Computer Programming

4.2.1 Qualitative analysis

For the qualitative analysis, which aimed at understanding students' perceptions regarding the use of ChatGPT in learning computer programming, semi-structured interviews were employed as the primary data collection method. This format granted flexibility, enabling participants to share detailed opinions and perceptions. At the same time, adherence to a set of predefined questions ensured consistency across all interviews. From the total sample of 26 students, 16 were chosen for these interviews. The 16 students were selected based on their levels of engagement with ChatGPT and their willingness to provide in-depth feedback, as assessed during the conduct of in-lab assignments 1 and 2. The remaining 10 students were not included in the interviews due to their limited availability for participation, which is caused by the lower levels of engagement with ChatGPT or scheduling conflicts. This selection ensured a diverse range of experiences and perspectives were represented, especially concerning the utilization of ChatGPT to complete assignments, homework, and take-home exams. Table 5 shows the examples of codes, theme, along with some students quotes collected from the interviews.

Table 5 Themes, codes, and examples of student responses

Theme	Codes	Examples from Student Responses
Impact of ChatGPT on Student Performance	Performance improvement, exam outcomes, lab assignments, performance variation	"Using ChatGPT helped me quickly understand the problem during the exam, which improved my performance."
Context Dependency of ChatGPT's Effectiveness	General vs. specialized queries, resource availability, exam type	"For general questions, ChatGPT was really helpful, but it didn't provide advantage for the specialized questions in our second exam."
Limitations and Drawbacks of ChatGPT	Limitations, student challenges, information management	"ChatGPT provides huge information about programming when I'm prompting it. It's not easy to use it in certain situations."
Comparison of ChatGPT to Traditional Learning Methods	Traditional vs. AI-based learning, integration with existing tools	"We need textbooks, online tutorials, and instructor-led classes to be prepared for midterm exams and final exams."
Integration of ChatGPT into Educational Curricula	Curriculum design, educational technology, AI integration strategies	"We need to use ChatGPT in learning programming as it gives immediate and direct response, but we need to monitor its use to prevent cheating."

4.2.2 Students' experiences and perceptions of ChatGPT in programming education

To understand the role of ChatGPT in computer programming education, several key themes and patterns emerged from the semi-structured interviews. The qualitative analysis highlights a spectrum of student experiences, ranging from the tasks and challenges they faced, the help they got from ChatGPT, the limitations of ChatGPT, and how ChatGPT compares to traditional educational methods.

Impact of ChatGPT on Student Performance:

For many students interviewed, the served as a supportive tool, especially when solving assignments, writing reports, and searching for appropriate data structures to save the data used in developed programs. One student shared, “I am using ChatGPT frequently because it gives me the proper data structures like dictionary, list, and tuple to be used in saving my data.” This reflects the platform's capacity to provide immediate and actionable insights that can streamline the coding process.

Context Dependency of ChatGPT's Effectiveness:

In regard to the tasks and challenges, students found ChatGPT useful in guiding them through the programming libraries. One student noted, “I usually use the ChatGPT to know which library I should use to complete the program because the programming language has many libraries.” This indicates a reliance on ChatGPT for navigating the language libraries, which can often be overwhelming for new students. Some other common challenges in the responses are explaining code, searching for the appropriate syntax, identifying the right libraries to import, and pinpointing runtime errors in their programs.

Limitations and Drawbacks of ChatGPT:

However, the tool wasn't without its limitations. A student remarked, "ChatGPT provides huge information about the programming when I'm prompting it. It's not easy to use it in certain situations." This highlights a potential issue where the information overload can make it challenging to quickly extract specific solutions.

Comparison of ChatGPT to Traditional Learning Methods:

The interviews also provided comparison of ChatGPT with traditional learning tools. While students saw value in ChatGPT for specific tasks, they emphasized the essential role of textbooks, online tutorials, and instructor-led classes. "We need textbooks, online tutorials, and instructor-led classes to prepare ourselves for midterm exams and final exams," commented one student. Another student added, "The ChatGPT is useful in preparing ourselves for exams because we use it to solve examples provided by the instructor in the PPTs and textbooks." These quotes reflect that while ChatGPT is a valuable asset, it serves as a complement rather than a replacement for traditional learning tools.

Integration of ChatGPT into Educational Curricula:

Regarding the potential integration of ChatGPT into educational curricula, there was a clear feeling towards its inclusion. "We need to use the ChatGPT in learning programming as it gives immediate and direct response," shared a student. However, some expressed difficulties, emphasizing the importance of oversight: "We need these emerging technologies like AI in teaching and learning, but we need to monitor the use of it to prevent cheating." This highlights the balancing required to ensure that while students benefit from advanced tools, the integrity of the learning process should be considered.

4.2.3 Response to the research question: What are the students' perceptions regarding the use of ChatGPT in learning computer programming?

Based on the qualitative findings from the semi-structured interviews, students' perceptions regarding the use of ChatGPT in learning computer programming are generally positive. They considered ChatGPT as a valuable asset in learning computer programming. It offers real-time guidance and support. However, they also emphasized the importance of its role as a supplementary tool, working alongside traditional teaching methods, rather than replacing them. The primary sentiment was that while ChatGPT can enhance the learning experience, it should be used wisely, with the appropriate checks in place to ensure academic integrity.

4.3 The Effective Method for Integrating ChatGPT into a Computer Programming Curricula

Based on both quantitative and qualitative data analyses, the effective approach for integrating ChatGPT into a computer programming curricula in determined in this section. The mixed methodologies followed in this study helped provide a comprehensive understanding of how ChatGPT can be integrated without compromising the core essence of traditional teaching. The detailed clarification of this integration strategy is presented below.

4.3.1 Response to the research question: What is the effective method for integrating ChatGPT into a computer programming curricula?

The answer to these questions, stemming from both quantitative and qualitative data analyses, requires a blend of state-of-the-art tools and the traditional pedagogies. ChatGPT, despite its impressive capabilities, should not replace the foundational instructional methods. Instead, it should serve as an additive component, amplifying the learning environment. The challenge in integrating ChatGPT lies in finding the right balance: granting students access to ChatGPT while keeping the integrity. This challenge becomes evident in scenarios such as exams and assignments, where there's a possibility for unsupervised use of ChatGPT, moving students away from the traditional oversight of human proctors or advanced monitoring systems. The solution to this challenge is designing these assessments using localized or customized problems. Such an approach ensures that learning maintains its integrity, even as students utilize the power of contemporary tools like ChatGPT.

5. Conclusion

This study examined the impact of ChatGPT on computer programming students' performance in in-class exams and in-lab assignments. It also explored the computer programming students' perceptions regarding the interaction of ChatGPT into the computer programming curricula. Furthermore, the research aimed to identify the effective manner to integrate ChatGPT into computer programming curricula as a supplementary learning tool. In examining the implications of integrating ChatGPT into computer programming instruction, the study employed a mixed-methods, qualitative and quantitative, strategy. A quasi-experimental design was applied, contrasting an experimental group using ChatGPT with a control group that did not.

The results of the quantitative analysis showed that the integration of ChatGPT into the learning process has a pronounced impact on students' performance. The experimental group, which utilized ChatGPT, consistently displayed superior performance in both in-class exams and in-lab assignments when compared to the control group. This underscores the potential of ChatGPT as a valuable supportive learning tool in computer programming education. The superior scores of the experimental group suggest that ChatGPT may have facilitated a deeper understanding of computer programming concepts. The platform's ability to provide immediate feedback, clarify doubts, and present information in a user-friendly manner might have contributed to this enhanced understanding (Grassini, 2023).

Qualitative feedback from the semi-structured interviews revealed that students predominantly viewed the integration of ChatGPT in computer programming education positively. They appreciated the real-time guidance it provided, underscoring its merit as a supplementary tool rather than a substitute for conventional teaching methods (Shoufan, 2023).

An effective method to integrate ChatGPT into a computer programming curricula could be concluded from quantitative and qualitative analyses. Initially, it is important to set a balance between using modern tools and traditional teaching methods. While ChatGPT can amplify the learning environment, it should not replace foundational instructional techniques. The primary challenge is in situations like exams and assignments, where there is potential for unsupervised usage of ChatGPT. To address this challenge, assessments should be designed with localized or customized problems. This ensures that the learning process retains its integrity, even when students leverage advanced tools like ChatGPT.

The noticeable improvement in the performance of students who used ChatGPT suggests that the integration of such AI-driven platforms can lead to more effective learning outcomes. Institutions and educators should consider incorporating ChatGPT and similar tools to capitalize on their potential to facilitate a deeper understanding of complex subjects like computer programming. The positive feedback from students regarding ChatGPT's integration underscores its merit, but it also highlights the need for a balanced educational approach. It's crucial that while AI tools are integrated, foundational teaching methods are preserved. Furthermore, to maintain the integrity of assessments, educators should prioritize crafting localized or customized problems especially when the potential for unsupervised tool usage exists.

While the current study provides insights into the benefits of ChatGPT, there's potential for further research. This could explore the long-term impacts of using ChatGPT, its effectiveness at different stages in the study program, and its utility in other domains beyond computer programming.

References

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A. & Mohammed, A. A. (2023). Impact of ChatGPT on Learning Motivation: Teachers and Students' Voices. Journal of English Studies in Arabia Felix, 2(1), 41-49. https://doi.org/10.56540/jesaf.v2i1.51

Abramovich, S. (2021). Using Wolfram Alpha with elementary teacher candidates: From more than one correct answer to more than one correct solution. Mathematics, 9(17), 2112. https://doi.org/10.1080/0020739X.2010.493241

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. Journal of AI, 7 (1), 52-62. Retrieved from https://dergipark.org.tr/en/pub/jai/issue/77844/1337500 https://doi.org/10.61969/jai.1337500

Bengtsson, L. (2019). Take-home exams in higher education: A systematic review. Education Sciences, 9(4), 267.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 1–12. https://doi.org/10.1080/14703297.2023.2190148

Erümit, Ali Kürşat. "Effects of different teaching approaches on programming skills." Education and Information Technologies 25, no. 2 (2020): 1013-1037. https://doi.org/10.1007/s10639-019-10010-8

Erhel, S. and Jamet, E. (2013) 'Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness'. Computers & education, 67, pp. 156–167. https://doi.org/10.1016/j.compedu.2013.02.019

Grassini, S. (2023). Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings. Education Sciences, 13(7), 692. https://doi.org/10.3390/educsci13070692

Goksel, N., & Bozkurt, A. (2019). "Artificial intelligence in education: Current insights and future perspectives". In Handbook of Research on Learning in the Age of Transhumanism, pp. 224–236. IGI Global. https://doi.org/10.4018/978-1-5225-8431-5.ch014

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. BenchCouncil Transactions on Benchmarks. Standards and Evaluations, 2(4), 100089. https://doi.org/10.1016/j.tbench.2023.100089

Hassani, H. and Silva, E. S. (2023). The role of ChatGPT in data science: how AI-assisted conversational interfaces are revolutionizing the field. Big data and cognitive computing, 7(2), 62-78. https://doi.org/10.3390/bdcc7020062

Joshi, I., Budhiraja, R., Dev, H., Kadia, J., Ataullah, M. O., Mitra, S., ... & Akolekar, H. D. (2023). ChatGPT--A blessing or a curse for undergraduate computer science students and instructors. arXiv preprint arXiv:2304.14993. https://doi.org/10.48550/arXiv.2304.14993

Kanika, Chakraverty, S., & Chakraborty, P. (2020). Tools and techniques for teaching computer programming: A review. Journal of Educational Technology Systems, 49(2), 170-198. https://doi.org/10.1177/0047239520926971

Lertbanjongngam, S., Chinthanet, B., Ishio, T., Kula, R. G., Leelaprute, P., Manaskasemsak, B., ... & Matsumoto, K. (2022, October). An empirical evaluation of competitive programming ai: A case study of alphacode. In 2022 IEEE 16th International Workshop on Software Clones (IWSC) (pp. 10-15). IEEE. https://doi.org/10.1109/IWSC55060.2022.00010

National Academies of Sciences, Engineering, and Medicine (2018). Appendix C: Commissioned Paper: Workforce Trends in Computer Science. In Assessing and Responding to the Growth of Computer Science Undergraduate Enrollments. Washington, DC: The National Academies Press. https://doi.org/10.17226/24926

Qureshi, B. (2023). Exploring the use of ChatGPT as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges. arXiv preprint arXiv:2304.11214. https://doi.org/10.48550/arXiv.2304.11214

Rosenberg, M. (2023, January 16). AI Writing Programs Raise Questions About Cheating in Universities. The New York Times. https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html

Royston, P. (1992). Approximating the Shapiro-Wilk W-test for non-normality. Statistics and computing, 2, pp. 117–119. https://doi.org/10.1007/BF01891203

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching, 6(1), 1-22. https://doi.org/10.37074/jalt.2023.6.1.9

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns, Healthcare, 11(6), 887. https://doi.org/10.3390/healthcare11060887

Sawilowsky, S. S. (2005) 'Misconceptions leading to choosing the t test over the Wilcoxon Mann-Whitney test for shift in location parameter'. Journal of Modern Applied Statistical Methods, 4(2), 26. https://doi.org/10.22237/jmasm/1130804700

Shakeel, A., Shazli, T., Salman, M. S., Naqvi, H. R., Ahmad, N., & Ali, N. (2021). Challenges of unrestricted assignment-based examinations (ABE) and restricted open-book examinations (OBE) during the COVID-19 pandemic in India: An experimental comparison. Human Behavior and Emerging Technologies, 3(5), 1050–1066. https://doi.org/10.1002/hbe2.290

Şenel, S., & Şenel, H. C. (2021). Use of take-home exam for remote assessment: A case study from Turkey. Journal of Educational Technology and Online Learning, 4(2), 236–255. https://doi.org/10.31681/jetol.912965

Spiegel, T. and Nivette, A. (2023. The relative impact of in-class closed-book versus take-home open-book examination type on academic performance, student knowledge retention and wellbeing. Assessment & Evaluation in Higher Education, 48(1), 27–40. https://doi.org/10.1080/02602938.2021.2016607

Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. IEEE Access.

Tikva, C., & Tambouris, E. (2021). A systematic mapping study on teaching and learning Computational Thinking through programming in higher education. Thinking Skills and Creativity, 41, 100849. https://doi.org/10.1016/j.tsc.2021.100849

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). 'User Acceptance of Information Technology: Toward a Unified View'. MIS Quarterly, 27(3), 425-478. https://doi.org/10.1109/ACCESS.2023.3268224

Wang, Y., Li, H., Feng, Y., Jiang, Y., & Liu, Y. (2012). Assessment of programming language learning based on peer code review model: Implementation and experience report. Computers & Education, 59(2), pp. 412–422. https://doi.org/10.1016/j.compedu.2012.01.007

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. Education and Information Technologies, 1-25. https://doi.org/10.1007/s10639-023-11742-4

Zielinski, C., Winker, M., Aggarwal, R., Ferris, L., Heinemann, M., Lapeña, J. F., ... & Citrome, L. (2023). Chatbots, ChatGPT, and scholarly manuscripts-WAME recommendations on ChatGPT and chatbots in relation to scholarly publications. Afro-Egyptian Journal of Infectious and Endemic Diseases, 13(1), 75–79. https://doi.org/10.21608/AEJI.2023.282936

EJEL Editorial 2024: Special Issue on AI in Education: Opportunities and Challenges (Parts 1 & 2)

Paul Griffiths[1] and Marija Cubric[2]
[1]EM-Normadie Business School/Metis Lab, UK
[2]University of Hertfordshire, UK

pgriffiths@em-normandie.co.uk (corresponding author)
m.cubric@herts.ac.uk

https://doi.org/10.34190/ejel.22.6.3831

The advent of Chat GPT and the other GenAIs that followed provides a very clear impression of the impact that machine learning will have on society. The impact of machine learning had been discussed for more than a decade, but those discussions had a high degree of abstraction and temporal speculation. The irruption of GenAI gives the discussion tangibility and a sense of urgency. It is now essential to define how society is going to govern this technology but before we do that we need to understand the impact that the technology is going to have on the different dimensions of society. In this special issue we attempt to do precisely that in the area of education.

The aim of this Special Issue is to bring together the latest research and developments in AI in education, to better understand how AI can enhance teaching and learning, and to identify the challenges and opportunities in this field. While the first reaction to ChatGPT in academic circles was one of seeing it as CheatGPT, the papers herein demonstrates that we have swiftly moved well beyond that initial state. The papers in this 2-part special issue are depicted in Table 1.

Table 1: Papers in Part 1 and Part 2 of the EJEL Special Issue on AI in Education

Paper id	Citation
PART 1	
#1.1	Cronje, Exploring the Role of ChatGPT as a Peer Coach for Developing Research Proposals: Feedback Quality, Prompts, and Student Reflection
#1.2	Humble et al., Cheaters or AI-Enhanced Learners: Consequences of ChatGPT for Programming Education
#1.3	Klyshbekova, and Abbott, ChatGPT and Assessment in Higher Education: A Magic Wand or a Disruptor?
#1.4	Watanabe, Have Courage to Use your Own Mind, with or without AI: The Relevance of Kant's Enlightenment to Higher Education in the Age of Artificial Intelligence
#1.5	Jose, and Jose, Educators' Academic Insights on Artificial Intelligence: Challenges and Opportunities
#1.6	Tseng, and Lin Enhancing English as a Foreign Language (EFL) Learners' Writing with ChatGPT: A University-Level Course Design
#1.7	Naz, and Robertson, Exploring the Feasibility and Efficacy of ChatGPT3 for Personalized Feedback in Teaching
#1.8	Alrayes, Henari, and Ahmed, ChatGPT in Education – Understanding the Bahraini Academics Perspective
PART 2	
#2.1	Hmoud et al.,

Cite this article: Griffiths, P. and Cubric, M. 2024. "EJEL Editorial 2024: Special Issue on AI in Education: Opportunities and Challenges (Parts 1 & 2)", *Electronic Journal of e-Learning*, 22(6), pp 82-87, https://doi.org/10.34190/ejel.22.6.3831

Paper id	Citation
	SI AI Rubric Development and Validation for Assessing Tasks' Solving via AI Cha
#2.2	Pingmuang et al., SI AI Exploring Generative-AI as a Learning Assistant - Understanding Thai Higher Education Students' Adoption: Technology Acceptance Model Research
#2.3	Håkansson Lindqvist et al., SI AI Exploring student and AI generated texts. : Reflections on reflection texts.
#2.4	Castelló-Sirvent et al., SI AI Quo Vadis, University? Discussing a Roadmap on Artificial Intelligence, Ethics
#2.5	Schaeffer et al., SI AI Risks of AI Applications Used in Higher Education
#2.6	Abdulla et al., SI AI Using ChatGPT in Teaching Computer Programming

In Cronjé's paper (#1.1) ChatGPT is used as a coach to help inexperienced researchers overcome the ambiguity that characterises the early stages of defining a research proposal. It is an adaptation to ChatGPT of what is a human-to-human peer coaching method known as GROW (which is an acronym for Goal, Reality, Opportunities, Will.) The paper describes an archival desk-study that analysed a sample of four worksheets out of a population of 93, each one produced by a student that took part of the research methods course. The sample was selected purposely based on that the worksheets "exhibited the clearest characteristics of the phenomenon under study." The students all started with a common prompt to kick off what then became a flow of interactions with ChatGPT. The students were encouraged to reflect on ChatGPT's responses, and at the end of the exercise on their experience with the tool throughout the exercise. Three clear issues emerge from the research: the quality of the feedback given by ChatGPT; the value of the prompts; and the importance of the student's reflection. The author ends with a useful recommendation that novice students given training on how to develop meaningful prompts, react to GenAI feedback, and engage in critical reflection .

Humble et al. (#1.2) set out to enquire on the potential consequences of ChatGPT on programming education in Computer Science. They apply a research method inspired by analytic autoethnography to understand the impact of this technology through the personal experiences of the authors. Their thematic analysis (of six field diaries, 82 interactions with ChatGPT and additional reflection notes) helps them, on the one hand, understand the strengths and weaknesses of ChatGPT that teachers should keep in mind, such as importance of quality of questions (prompts), factual (textbook knowledge) excellence, 'sloppiness'. And, on the other hand, understand the potential consequences that using ChatGPT could have on teaching. ChatGPT can be an obstruction to learning or a great assistant to teachers and students and as a consequence will require re-thinking teaching and learning practices. In short, Computer Science students are going to use GenAI; it is now a question of seeing whether they do it alone with the risks that that entails, or whether teachers take a lead in shaping the future.

Klyshbekova & Abbot (#1.3) set out to do an experiment on the capabilities of ChatGPT-3 to write and assess an essay. They use the Christensen et al (2015) definition of disruptive innovation to assess whether the use of ChatGPT-3 for this purpose is actually disruptive. They commend the GenAI tool for its speed in producing an essay and its human-like interactive style with the user in the process of production. They are not, however, impressed with the quality of the text in the sense that it shows limited creativity and critical thinking. They judge the essay dull as if producing filler-text and highly unreliable in its use of academic references. The experimental part of the work consists in getting ChatGPT-3 to write an essay, define a marking guide, and then mark its own work. The researchers then created their own marking guide following a well-established set of criteria, and used it to assess ChatGPT's essay. They arrive at the human-like condition that ChatGPT rates its own work far higher than do its human appraisers! Is ChatGPT developing consciousness, subjectivity, self-interest? On a more serious note, this paper leaves the reader thinking: Given that ChatGPT consistently gives a different outcome as response to the same prompt, can it ever be trusted to give fair assessments? Furthermore, the authors are calling for revisiting current assessment practices that may be vulnerable to cheating from AI-generated content.

Watanabe (#1.4) does a theoretical study deeply rooted in philosophy (ethics) to assess the effects of AI on education. It relates the use of AI in universities to Kant's reflections on enlightenment. The explicit aim of the article is to theoretically analyse the compatibility of several AI tools with the ideal of maturity on an educational philosophical level. Maturity has many dimensions to it, but the author opts to look at it from the perspective of critical thinking and independent action. The findings can be categorised into three areas. The first is that Intelligent Tutoring Systems (ITS) deprive the students of independent assessment and reflection on their learning. Moreover, students cannot discuss ITS suggestions or assessments, which discourages critical enquiry. Furthermore, it incentivises students to focus on their own success and thus impairs collaboration and teamwork. The second is that ChatGPT limits students' writing and reading abilities and therefore negatively affect the students' competences for independent interpretation of texts. It leads to passiveness and a lack of independent thinking. Finally, only AI-based research and visualisation tools lead students to conduct independent research without excessive dependence on AI, which promotes development of maturity.

"What are educators' perceptions of the challenges and opportunities of using AI tools in learning and teaching?" is the research question addressed by Jose & Jose (#1.5). They do a qualitative study with a sample of 35 educators in the English language, belonging to two networks at the University of Technology and Applied Science Al Musannah and at the English Language Exchange, Oxford University. They arrive at three main themes: Concerns, Risks/Challenges and Opportunities. If numbers mean anything, it is telling that under Concerns they arrive at eleven sub-themes, ten subthemes for Risks/Challenges and a massive 22 for Opportunities. A basis for optimism? Much of the Concerns arise from AI technology, its rapid rise and possibly hyped popularity. In terms of Risks, the educators worry about job displacements, the stifling of critical thinking, giving feedback on AI-assisted writing and the development of laziness. Finally, the Opportunities: providing templates for writing, learner motivation, AI as an educational aid, its use in teaching potentially dangerous subjects, delivering personal learning experiences, providing interactive tutoring and several others.

Continuing with English as a Foreign Language at University level, only that in this case it is a writing course in Taiwan, Tseng & Lin (#1.6) combine an instructional design model (ADDIE) with a technology-enabled pedagogical model (TPACK) to propose an integrated writing course framework that they test with a sample of 15 students. Primary data collection is done through two stages, the analysis of the students' written work followed by their reflective writings. The authors give a positive outlook stating that the students developed an ethical authorship conscience and a critical stance to combine their ideas with the output of ChatGPT. In their findings the authors reveal that ChatGPT addresses three fundamental challenges often encountered in academic writing courses: (a) it enhances efficiency by by providing immediate feedback and generating content ideas; (b) it helps achieve a cohesive organisation within students' writing, guiding them to structure their thoughts more logically; and (c) the students declared in their reflective pieces that ChatGPT helped them eliminate errors by giving them objective feedback that they used to refine their drafts. In conclusion, ChatGPT can effectively be integrated into writing instruction in a form that does not create dependence as long as it is done through a structured and pedagogical framework.

In the final case of language learning, Naz & Robertson (#1.7) explore the efficacy of ChatGPT feedback in written English learning. The authors underpin their work on social cognitive theory (SCT) which supports the use of GenAI in that it gives instant guidance and support that enables personalised, independent learning; and on second language acquisition theory (SLA) that supports that the use of GenAI can enhance student learning by providing meaningful interaction. They use a small sample of four written pieces from four different students, of different lengths and content matter. The authors design a rubric and mark the pieces via ChatGPT and humans using the same rubric, and compare the results. The authors find that ChatGPT can be a great tool through giving students timely feedback and also interaction and motivation. There are problems of errors and hallucinations, but this can be minimised by putting care into the design of the marking guide and complementing it with human oversight. However, the outcome is that ChatGPT can be very helpful but there is some critique on the vagueness of feedback, especially on technical subjects. A key finding is that, the more precise the marking criteria, the more reliable the outcome.

Alrayes (#1.8) take a completely different approach in that they look at the problem of adopting ChatGPT in higher education from the perspective of the educator and in that they do a quantitative study. They address the issues of how socio-demographic factors influence the adoption of ChatGPT by educators; what motivates them and how it impacts their teaching practices; and what social influences affect academics in using ChatGPT. They apply the Unified Theory of Acceptance and Use of Technology (ATAUT) model but reflecting the fact that higher education institutions have not adopted policies on using GenAI tools and therefore it is voluntary for the academics to adopt it, they apply the UTAUT2 version for consumers. Their survey received 141 responses of which the majority are Millennials or Generation X, are highly educated, with a predominance of women, of whom less than half have used ChatGPT or any other GenAI tool. They conclude that education institutions must embark on a strategic roadmap not only for the adoption of GenAI but also adapt to continuous advancements in the technology. They recommend training both educators and students on their use, and make considerations on ethical aspects. Two of the ethical precautions they make are commendable but unpractical, as they propose transparency in the sense of explicability of how ChatGPT arrives at its responses, and concerns around student data collection, both of which are intrinsic to how current large learning models work.

In their study, Hmoud et al. (#2.1) aimed to systematically evaluate the educational impact of AI by creating and testing a rubric designed to assess student work when supported by AI chatbots. The process began with a review of the literature, from which the authors identified 37 key assessment criteria. These were evaluated by a panel of 12 ICT professionals and applied to the final assignments of 144 education students. After applying various statistical methods for content validation, the list was refined to 9 essential items, which were grouped into two categories based on factor analysis. The first, "Quality of Content," encompassed criteria like accuracy, relevance, coherence, comprehensiveness, grammar and spelling, argumentation and evidence. The second, "Quality of Expression," included efficiency, language and tone, creativity and originality. While this research contributes to the fast-developing area of AI chatbot assessment, the authors acknowledge that further validation is needed—particularly with larger sample sizes and in more varied educational contexts.

Kanont et al. (#2.2) used the Technology Acceptance Model (TAM) to explore factors influencing the adoption of GenAI tools among Thai university students. Surveying 911 students across various subjects, they found that Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) significantly impacted adoption. Interestingly, PU and PEOU were negatively correlated, suggesting that as GenAI becomes more intuitive, students may see it as less valuable for learning. This highlights a paradox where students, while open to GenAI, recognise potential downsides to its educational impact. The authors also discuss the importance of universities promoting ethical AI integration, and address challenges like language barriers for non-native English speakers. They also suggest future research should focus on qualitative data and longitudinal studies to gain deeper understanding of the factors influencing adoption and how they evolve over time.

Castelló-Sirvent et al.'s bibliometric study (#2.3) examines AI and academic integrity in higher education (HE), analyzing 254 studies from high-impact journals. The research identifies four key themes: i) technological and design factors influencing AI adoption in HE , ii) machine learning-based prediction of student performance, iii) AI's role in interventions affecting academic integrity and student attitudes, and iv) constraints in the learning environment that AI can help address. The study offers practical insights for educators, academic authorities, and policymakers, highlighting issues such as ethical AI use, faculty commitment to academic integrity, AI's financial and organisational impact, and its influence on social mobility.

Håkansson Lindqvist and Arvidsson's study (#2.4) compares students' reflections (N=20) with ChatGPT-generated texts, revealing notable content differences. While students provided reflective texts with in-depth, personal insights into their learning, ChatGPT offered only questions to guide and support reflection. Thus, the AI did not produce texts that students could submit (i.e., plagiarise); instead, it helped guide students toward deeper reflections. These findings may assist teachers in designing assessments, and the study could be expanded to a larger sample and exploring different types of reflection.

A study by Abdulla et al. (#2.5) seeks to integrate ChatGPT into computer programming curricula to enhance students' learning. The research employs a quasi-experimental design, comparing an experimental group that used ChatGPT with a control group that did not. A total of 26 undergraduate students from a university in Oman participated. Evaluation methods included two in-class exams and two in-lab tasks. Quantitative analysis showed that while ChatGPT can be a valuable tool for general programming queries, its effectiveness may vary with highly specific questions. The experiment was further supported by qualitative data from semi-structured interviews with 16 participants from the same group, revealing that students found ChatGPT to be a valuable learning aid and that combining ChatGPT with traditional teaching methods could further optimise learning outcomes. The authors recommend further research to explore the long-term effects of AI integration in educational settings.

Schaeffer et al.'s viewpoint article (#2.6) examines the risks of AI applications in higher education and aims to initiate discussion among developers, decision-makers, and users about mitigating these risks. The main concerns include transparency of AI models, user privacy, data confidentiality, integrity, system availability, and bias, all stemming from data-related issues such as data being used beyond its intended purpose or collected from unintended sources. The authors recommend adopting a recognised risk management framework to address these challenges and conclude that the benefits of AI in higher education can only be realised if institutions enforce strict governance and compliance measures.

Reflecting on the content of the papers included in this special issue we extract and synthesises the key concepts from these papers, which we show in Table 2.

Table 2: Key concepts that arise from the papers in the Special Issue, Parts 1 &2 (the numerals in the reference columns correspond to the numerals in Table 1)

Used correctly GenAI has strengths:	References	Used incorrectly GenAI has weaknesses:	References
EFFICIENCY 1.Speed to produce essays 2.Generates content ideas	#1.3 #1.5, #1.6		
INTERACTIVE FLOW 3.Provides immediate feedback 4.Performs human-friendly interaction 5.Provides interactive tutoring 6.Provides Peer-to-Peer like coaching to inexperienced researchers 7.Assisting students in task-solving e.g. drafting outlines, revising content, proofreading, and post-writing reflection	#1.6 #1.3 #1.5, #1.7 #1.1 #2.1, 2.4		
RELIABILITY 8.Weakness 3. Can be minimised with good design of rubric and complemented with human oversight	#1.7, #2.1	1.Unreliable use of academic references 2.Hallucinations 3.Given that the same prompt can different outcomes: Can it be trusted for student assessment? 4.Inaccuracy & incompleteness	#1.3 #1.1, #1.7 Reflection on #1.3 #2.6
EDUCATIONAL AID 9.Leads to motivation 10.Useful for teaching potentially dangerous subjects 11.Enables personalised learning 12.Can support task-solving assessments when evaluated against the identified criteria. 13.Can support reflection	#1.5, #1.7 #1.5 #1.5 #2.1 #2.4	5. Educators worry about job displacements 6.Vagueness of feedback	#1.5 #1.7
QUALITY OF WRITING 14.Cohesive structure of written reports	#1.6	7.Poor quality of text - No creativity - Dull, filler text	#1.3
STUDENT DEVELOPMENT & MATURITY 15.AI-based research assistance and data visualisation lead to independent thinking 16.Develops increased authorship conscience 17.Increases critical stance from combining own ideas with GenAI output 18.Done within pedagogical framework does not create dependency	#1.1, #1.4 #1.6	8.Use leads to poor critical thinking 9.Inhibits independent action 10.Incentivises students to focus on their own success and not teamwork 11.Limits development of reading and writing skills	#1.3, #1.4, #1.5 #1.4 #1.4

Used correctly GenAI has strengths:	**References**	**Used incorrectly GenAI has weaknesses:**	**References**
	#1.6 #1.6	12.Creates excessive dependency 13.Development of laziness 14. Students' privacy & ethical concerns	#1.4 #1.4 #1.5 #1.8, #2.2
STUDENT PERFORMANCE 19. Increased student performance in general (subject) knowledge and less so in more specific questions	#2.5	15.Negative impact on learning	#2.2

In summary, the findings from these papers suggest that the higher education community is cautiously optimistic about the potential of AI tools, while also being aware of the associated risks and pitfalls (Pelletier et al., 2024). From our analysis, we derive the following insights:

- It is clear that students are already using GenAI and will continue doing so. So, are we going to leave them to use it on their own, without guidance, with all the risks that that entails? Or will educators take leadership and guide them? (#1.2)
- The advent of machine learning and their adoption by higher education institutions will require a re-think of teaching, learning and assessment practices. (#1.3)
- Higher education institutions need to embark in incorporating GenAI and its future development in their strategic roadmap.(#2.2, #2.3)
- Higher education institutions need to define ethical guidelines for the use of GenAI, and train educators and students on their use and on understanding of the ethical implications of their use. (#1.8, #2.1, #2.2, #2.3)
- When properly assessed and integrated AI chatbots could be valuable in higher education (#2.1, #2.4, #2.6).

We are convinced that this special issue meets its objective of giving students and educators a better understanding of how GenAI can enhance teaching and learning, as well as help lay down some foundations for the governance of GenAI and its successor machine-learning solutions in the educational domain.

References

Pelletier, K., Mccormack, Muscanell, N, Reeves, J., Robert, J., Arbino, N., (2024). *2024 EDUCAUSE Horizon Report: teaching and learning edition* (pp. 1-46). EDUC24.