

EVALUATING THE PERFORMANCE OF CHATGPT IN ACCOUNTING AND AUDITING EXAMS: AN EXPERIMENTAL STUDY IN NORTH MACEDONIA

Atanasko Atanasovski

Full Professor at Ss. Cyril and Methodius University, Faculty of Economics – Skopje
atanasko@eccf.ukim.edu.mk

Todor Tocev

Assistant at Ss. Cyril and Methodius University, Faculty of Economics – Skopje
todor.tocev@eccf.ukim.edu.mk

Ivan Dionisijev

Assistant at Ss. Cyril and Methodius University, Faculty of Economics – Skopje
ivan.d@eccf.ukim.edu.mk

Zoran Minovski

Full Professor at Ss. Cyril and Methodius University, Faculty of Economics – Skopje
zminovski@eccf.ukim.edu.mk

Dimitar Jovevski

Associate Professor at Ss. Cyril and Methodius University, Faculty of Economics – Skopje
dimitar.jovevski@eccf.ukim.edu.mk

ABSTRACT

In the ever-evolving landscape of education, the integration of artificial intelligence (AI) has emerged as a transformative force, reshaping conventional pedagogical methods and assessment approaches. The domains of accounting and auditing, integral for preparing future professionals, have not remained immune to this technological revolution. As AI, including ChatGPT, gains prominence, it offers the potential to enhance learning and examination experiences within these disciplines.

This paper presents a pioneering North Macedonian experimental research evaluating the performance of ChatGPT's effectiveness in exam performance in accounting and auditing. Leveraging 11 subject exams comprising a total of 401 questions, this research assesses ChatGPT's proficiency in addressing intricate, context-specific questions within these domains.

ChatGPT 3.5 demonstrated its reliability as a "student" by successfully passing 8 out of the 11 subjects, achieving a pass rate of 73%. The findings reveal that ChatGPT 3.5 excels in responding to qualitative open-ended questions, true/false assessments, and multiple-choice questions with a single correct answer, effectively streamlining routine tasks and offering comprehensive explanations. However, it faces challenges when dealing with quantitative calculations and complex multiple-choice questions with multiple correct answers.

This study contributes to the ongoing discourse on AI's role in education, sparking discussions about student access to AI tools, the balance between memorization and referencing, the potential for AI to enhance learning experiences, and the unique contributions of educators that surpass AI capabilities. As AI continues to disrupt traditional practices across professions, including education, its responsible integration promises to enhance the educational process and deliver substantial value. ChatGPT, as a versatile AI language model, stands as a powerful tool in this

<http://doi.org/10.47063/EBTSF.2023.0003>

<http://hdl.handle.net/20.500.12188/28841>

endeavor, offering educators and students the potential to enrich their learning journeys while recognizing the need for judicious application within its defined constraints.

Keywords: *AI, ChatGPT, accounting and auditing, higher education*

JEL classification: *M40, I23, O33*

1. INTRODUCTION

In the rapidly evolving field of education, the incorporation of artificial intelligence (AI) has become a transformative force, redefining traditional teaching methods and evaluation strategies. Education in accounting and auditing, crucial for preparing future professionals, has not remained untouched by this AI technological surge. The accounting profession often perceived as a domain of numbers and financial data, has undergoing a profound transformation and expansion of the domain in analysis, interpretation, and providing support for informed decision-making. AI is emerging as a disruptive technology that is reshaping the way accountants work, manage and analyze data, and provide insights.

The incorporation of AI-driven tools, such as ChatGPT (Generative Pre-trained Transformer), offers a potential for enhancing the learning and examination experience in these disciplines.

ChatGPT is a large language-learning model that employs machine learning to produce human-like text in natural language (Susnjak 2022). It represents a "General Purpose Technology", as typically defined in broader economic or technological contexts (Wood et al., 2023).

When asked "What are you and how are you designed", the answer was as follows:

"I am ChatGPT, a computer program developed by OpenAI. I'm designed to understand and generate human-like text based on the input provided to me. My purpose is to assist with answering questions, providing information, generating text, and engaging in text-based conversations on a wide range of topics." He also emphasizes that he is a machine learning model and not a sentient being: "My responses are generated based on patterns in the data I was trained on, and I don't possess consciousness, emotions, or personal experiences".

About the way Chat GPT works indicates that it is trained in two stages, namely: (1) Pre-training phase: *"I learn from a vast amount of internet text data, focusing on predicting the next word in a sentence to grasp language patterns, grammar, facts, and common sense"*, and (2) Fine-tuning, which is after the pre-training phase: *"I'm refined for specific uses. I'm trained on a curated dataset with human reviewers following guidelines. This customizes my behavior, aligns it with specific applications, and ensures safety and appropriateness"*.

Large language models (LLMs) are a category of language models known for their exceptional performance in various natural language processing (NLP) tasks (DAIR.AI, 2023). They have gained significant attention in research due to their capacity to produce human-like language and their potential to bring about transformative advancements in science and technology.

Brownlee (2023) emphasized that AI chatbots, like ChatGPT, mark a significant departure from "traditional" search engines (e.g. Google), as they have the capacity to read a lot of pages and generate answers from reliable web content, a development poised to bring increased sophistication to the accounting profession, underscoring the importance of understanding how to effectively harness this technology.

Based on initial evaluations by both users and experts, ACPM (2023) observes that ChatGPT has the potential to assist accountants by automating various tasks in the future. These tasks encompass

tasks like formulating Excel, Power BI, Tableau, and similar formulas, transactional accounting, low-level computer programming, as well as the interpretation of specific accounting standards and valuation. Nonetheless, it's essential to note that any accounting data generated by Chat GPT should be subject to expert verification (ACPM, 2023).

This paper presents qualitative and experimental research on the performance of the ChatGPT chatbot "in the exam room". The research aims to perceive the performance in ChatGPT responses by conducting 11 subject exams i.e. 401 variously structured questions in total. These subjects are integral to the accredited Accounting and Auditing curriculum at the undergraduate level of the Faculty of Economics-Skopje, part of the Ss. "Cyril and Methodius" University in Skopje. ChatGPT's training encompasses a broad range of topics, and while it does contain information related to accounting and auditing, its performance in these specific fields can be suboptimal. This is because faculty professors and textbook authors usually design questions that require nuanced understanding from students, a level of understanding that AI algorithms may struggle to grasp. Consequently, this paper seeks to provide empirical evidence regarding ChatGPT's proficiency in tackling challenging and context-specific questions within the domains of accounting and auditing.

The paper is structured into five conceptual parts. In the introductory section, we provide a concise overview to enhance comprehension of ChatGPT chatbot 3.5 - an exceptionally fast-growing and extensively applied AI technology platform. In the literature review section, we endeavor to summarize the most pertinent and up-to-date findings concerning ChatGPT, with a particular focus on its role in education, specifically within the realm of accounting and auditing. Then, the methodology of the experimental research is presented, followed by the results and a discussion of the findings. We also candidly address any limitations encountered during the study and offer recommendations for future research endeavors. In the last section, we present a comprehensive summary of all the conclusions derived from the conducted research.

To our knowledge, this study represents a pioneering effort in North Macedonian higher education by evaluating ChatGPT's effectiveness in enhancing exam performance. Our findings have the potential to revolutionize how students approach and excel in these challenging subjects, making a significant contribution to the ongoing discussion about AI's role in education. Such experimental research should raise debates about the use of AI tools in higher education and develop a larger level discussion on the following questions, raised by Wood et al., (2023): In what ways should students be granted access to AI? Which material should be memorized versus referenced? Can engaging with AI improve students' learning experiences, and if so, how? What unique contributions do professors and accountants offer that go beyond the capabilities of AI?

Scholars are progressively embracing AI as a disruptive technology and analyzing its implications across various professions, including education. Consequently, it's merely a matter of when and how this technology will be harnessed, with a keen consideration of its pertinence and adaptability to enhance the educational process and deliver added value.

2. LITERATURE REVIEW

AI academic research has a diverse and extensive historical background, spanning various fields, which include education, and also accounting and auditing as professions/fields.

Beerbaum (2023) explores the transformative potential of generative AI technology in the field of accounting, particularly in automating tasks and data generation. A notable application of generative AI within accounting lies in enhancing financial statement analysis. By training on extensive financial datasets, generative AI algorithms acquire the ability to discern patterns and correlations among diverse financial metrics (Beerbaum, 2023). This proficiency empowers

accountants and financial analysts to detect anomalies, trends, and potential risks, thereby facilitating more informed and strategic decision-making processes.

Frey and Osborne (2017) assess the vulnerability of job roles to automation, highlighting a significant likelihood of AI taking over accounting positions. In this context, Wood et al. (2023) suggest that the AI chatbots' accuracy in responding to accounting assessment questions may have implications for how AI's capacity to substitute accountants in practical applications is perceived.

Due to the extensive incorporation of contemporary information technologies (IT) in the accounting, including big data, data analytics, AI, blockchain, cloud computing, and others, there has been a shift in market demands for accounting proficiency.

Numerous authors, including Huq, (2014), Faggella, (2020), Han et al., (2023), Yan et al., (2022), Al Ghatrifi et al., (2023), examine and engage in discussions regarding the utilization of these aforementioned disruptive technologies. They assert that AI, whether employed as a distinct tool or integrated technology, holds the potential to substantially enhance accounting procedures and the audit process. These authors underscore, either explicitly or implicitly, the substantial influence of AI on the accounting field and promote the importance of adapting educational systems to equip new generations in addressing this challenge.

Accounting education must evolve in tandem with these changing expectations, emphasizing the need for integrating IT advancements into curricula (Yan et al., 2022). The inclusion of the latest technologies within accounting programs becomes imperative to equip students with the required skills for a constantly evolving field of accounting (Al Ghatrifi et al., 2023).

ChatGPT, as AI tool, has not only garnered significant public attention since its launch in November 2022 but has also piqued academic interest (George & George, 2023, Mathew, 2023). The research findings of Elbanna & Armstrong (2023) suggest that ChatGPT can be efficiently incorporated into education to automate repetitive tasks, improve the overall student learning experience, and consequently enhance productivity and efficiency while promoting adaptive learning. Nevertheless, it is crucial to acknowledge and be mindful of ChatGPT's limitations, even with updates. These limitations encompass factual inaccuracies, the potential for promoting biases, limitations in comprehensive understanding, and concerns related to safety (Elbanna & Armstrong, 2023).

In the realm of education, there exist remarkable prospects as well as formidable obstacles, ushering in a swift transformation from conventional teaching approaches to intelligent ones (Tan, 2023). Tan (2023) in his research quotes Adam Stevens who cautioned against prioritizing scores over knowledge in the education system, emphasizing that while ChatGPT offers tremendous potential as an auxiliary tool, it can also pose threats. Standardizing its use to avoid student over-dependence, improving security, accuracy, and reliability, and exploring broader educational applications are key steps for ChatGPT's development, leading to more comprehensive evaluation and oversight measures as all aspects progress (Tan, 2023, Elbanna & Armstrong, 2023). It is noteworthy that, despite various controversies and ethical considerations, ChatGPT has garnered significant interest from academia, research, and industries within a remarkably brief timeframe (Ray, 2023). In a survey carried out by Constantz (2023), approximately 30% of the participants reported attempting to utilize ChatGPT in their workplace.

Wood et al. (2023) present valuable insights into ChatGPT's performance in accounting education, utilizing a vast dataset contributed by a diverse team of 327 coauthors from 186 educational institutions worldwide. Their study encompassed an impressive 25,817 questions drawn from 869 distinct class assessments, along with an additional 2,268 questions sourced from textbook test

banks, spanning topics like accounting information systems, auditing, financial accounting, management accounting, and tax.

The research conducted by Wood et al. (2023) yielded the following findings:

- ChatGPT's performance varies depending on the subject area being assessed. Notably, it exhibited comparatively stronger performance in assessments related to accounting information systems (AIS) and auditing when compared to assessments related to tax, financial accounting, and management accounting.
- ChatGPT displayed higher accuracy rates in answering true/false and multiple-choice questions, achieving rates of 68.7 percent and 59.5 percent, respectively.
- Conversely, ChatGPT encountered challenges when dealing with practice questions and providing short answers, achieving lower accuracy rates of 28.7 percent and 39.1 percent, respectively.

Eulerich et al. (2023) conducted an investigation into the potential of newly released ChatGPT models and their capabilities to successfully pass major accounting certification exams, including the CPA, CMA, CIA, and EA certification exams. Their findings indicate that the initial release of ChatGPT 3.5 was unable to achieve a passing score on any of these exams, with an average score of 53.1 percent across all assessments. However, through additional refinements and enhancements, including the transition to the ChatGPT 4 model, 10-shot training, and permitting the model to use reasoning and external resources, ChatGPT demonstrated significant improvement, achieving an average score of 85.1 percent across all exam sections and successfully passing all exams (Eulerich et al., 2023).

Bommarito et al. (2023) conducted a thorough assessment of professional knowledge worker readiness using the Uniform CPA Examination from AICPA as a benchmark. They used OpenAI's text-davinci-003 model and earlier GPT versions to evaluate a sample REG exam section and 200+ multiple-choice questions covering various domains. Here are the main findings:

- Text-davinci-003 had a 14.4% accuracy rate in the sample REG exam section, indicating limitations in quantitative reasoning with zero-shot prompts.
- Text-davinci-003 showed promising performance, approaching human-level abilities, particularly in Remembering & Understanding and Application skills. With optimal prompts and settings, it answered 57.6% of questions correctly, surpassing random guessing, and its top two responses were accurate 82.1% of the time, demonstrating strong non-entailment reasoning.
- Recent iterations of GPT-3, including text-davinci-003, significantly improved performance, with the correct answer rate increasing from 30% for text-davinci-001 to 57% for text-davinci-003.

These findings highlight the potential of large language models to enhance the quality and efficiency of future knowledge work, suggesting that when optimized appropriately, these models can match or even surpass human-level performance in specific professional fields.

Nunes et al. (2023) anticipate significant impacts of LMs on education, potentially leading to AI-powered tools for psychometric exam analysis. In this regard, Kolade et al. (2023) suggest several propositions:

- i. Chatbots can provide instant formative text feedback to learners.
- ii. Automated essay scoring systems can assess summative assessments, saving staff time.
- iii. Computerized adaptive feedback (CAF) with AI can offer personalized assistance, enhancing engagement and study habits.
- iv. AI-driven computerized adaptive testing can adapt to learners' competence and personalized learning for summative assessment.

v. Computer serious games provide unlimited formative feedback in real-life contexts. In the study conducted by Newton & Xiromeriti (2023), it was observed that older and free versions of ChatGPT, based on GPT3 and GPT3.5, displayed rather modest performance levels. These earlier iterations struggled to pass most exams and consistently underperformed when compared to the average student. In stark contrast, ChatGPT4 (currently paid chatbot product ChatGPT Plus) exhibited a remarkable improvement in performance, surpassing its predecessors by a significant margin of 25 percentage points (Newton & Xiromeriti, 2023). This notable enhancement allowed ChatGPT4 to successfully pass the majority of the exams assessed. These findings hold profound implications for the validity of current Multiple-Choice Question (MCQ)-based assessments within Higher Education. They challenge the efficacy of unproctored online exams as a reliable summative assessment method (Newton & Xiromeriti, 2023). Therefore, it is imperative to reconsider the appropriateness of such assessment methods in light of these new insights.

3. RESEARCH METHODOLOGY

The experimental research encompassed 11 compulsory subjects within the Bachelor's degree program in the Accounting and Auditing department at the Faculty of Economics – Skopje, Ss. Cyril and Methodius University in Skopje (<https://eccf.ukim.edu.mk/en/studies/first-cycle-studies/accounting-and-auditing-programme/accounting-and-auditing-first-cycle-study-program-2022-23/>). It is noteworthy that this program holds accreditation from the Association of Chartered Certified Accountants (ACCA), highlighting its relevance and importance in the accounting field. Notably, students in this program receive recognition for seven out of a total of nine ACCA courses during their academic journey, underscoring the program's significance. ChatGPT was tested on the following subjects (see Table 1):

Table 1: Accounting and Auditing Exams – Sample Research

1) Principles of Accounting	7) Governmental Accounting and accounting of non-profit organizations
2) Financial Accounting	8) Financial statements analysis
3) Corporate Financial Reporting	9) International accounting
4) Management Accounting 1	10) Auditing
5) Management Accounting 2	11) Internal auditing
6) Accounting Information Systems	

(Source: Author's text)

The experimental design entailed assessing each of the 11 subjects by employing ChatGPT 3.5 (free version), functioning akin to a student, to answer exam questions. The questions' structures vary, mirroring the diverse assessment methods employed in the program. These assessment methods include i) true/false questions, ii) multiple-choice questions (encompassing various correct answer options), iii) open-ended short questions, and iv) essay questions (case studies). The scoring process for these questions aligns with the manual evaluation criteria applied in standard student assessments. Partial scoring was implemented for open-ended short questions and essay questions (including case studies). In these cases, points were granted for partially solved tasks or partially answered questions. Conversely, for true/false questions and MCQs, points were exclusively awarded for completely correct answers, with no points given for incorrect responses.

Data collection primarily revolves around the design and administration of exams across the 11 subjects. The methodology applied in the research is consistent with the experiments conducted by Wood et al., (2023), Eulerich et al. (2023) and Bommarito et al. (2023).

This experiment involves a qualitative assessment of ChatGPT 3.5's readiness to tackle exams in the domain of accounting and auditing. It delves into an analysis of ChatGPT's capacity to respond to various question categories, shedding light on both its strengths and limitations as an AI tool. Additionally, the paper explores the potential and opportunities presented by ChatGPT and other AI tools/LLMs in enhancing education through interactive learning.

4. RESULTS AND DISCUSSION

ChatGPT 3.5, functioning as a freely accessible AI tool, demonstrated a commendable level of performance when taking on the role of a student in accounting and auditing subjects. Table 2 provides a summary of its responses to all exam questions. Our evaluation considered two key aspects: the number of correct answers provided and the points earned, aiming to gauge the depth and precision of its responses.

Table 2: Overall evaluation of ChatGPT 3.5 Performance in Accounting and Auditing Exams

Questions types:	No. of questions	ChatGPT overall response rate (correct answers)	ChatGPT overall response rate (grade points)
True/False questions - (T/F)	80	65%	65%
Multiple choice questions (MCQ) with 1 correct answer	114	72%	60%
Multiple choice questions (MCQ) with multiple correct answer (1-5)	120	48%	48%
Open-ended short questions - (SQ)	60	78%	78%
Essay questions (case studies) - (EQ)	27	7%*	55%
<i>Total</i>	<i>401</i>	<i>60%</i>	<i>57%</i>

*Fully correct answer

(Source: Author's calculation)

Notably, ChatGPT performed most proficiently in open-ended short questions while facing more challenges in multiple-choice questions (MCQ) with multiple correct answers (1-5). It is worth highlighting that ChatGPT provided fully correct responses to only 2 out of 27 essay questions (case studies), yet it addressed all questions partially correctly, offering detailed explanations and establishing logical connections. If partial scoring were not applied to essay questions (case studies), it would result in the lowest scores for this type of question.

Regarding MCQ questions with a single correct answer, ChatGPT's performance was more favorable compared to those with multiple correct answers. As partial scoring was not applied to the latter, no points were awarded when ChatGPT answered these questions partially correctly.

In sum, based on the compiled results, it can be concluded that ChatGPT provided correct responses to 60% of the questions, equivalent to 241 out of a total of 401 questions. However, when considering the weight and scoring of the questions, it earned 57%, equating to 570 out of a possible 1000 grading points.

Table 3 provides a comprehensive breakdown of information for each subject individually. The overall pass rate for the analyzed subjects stands at 73% (8 out of 11 subjects), with a passing grade defined as achieving a minimum score of 50%.

Table 3: Detailed evaluation of ChatGPT 3.5 Performance in Accounting and Auditing Exams

	T/F		MCQ (1)		MCQ (1-5)		SQ		EQ (CS)		Total score	
	Q	R %	Q	R %	Q	R %	Q	R %	Q	R %	R%	G
Principles of Accounting	10	60%	20	80%	/	/	/	/	2	92%	81,5%	9 (B)
Financial Accounting	/	/	12	67%	/	/	/	/	5	62%	64%	7 (D)
Corporate Financial Reporting	/	/	12	58%	/	/	/	/	3	46%	52%	6 (E)
Management Accounting 1	10	60%	20	86%	/	/	/	/	5	35%	63%	7 (D)
Management Accounting 2	10	50%	20	56%	/	/	/	/	5	35%	46%	5 (F)
Accounting Information Systems	10	70%	/	/	30	47%	/	/	/	/	52,5%	6 (E)
Governmental Accounting and accounting of non-profit organizations	10	50%	/	/	30	40%	/	/	/	/	42,5%	5 (F)
Financial statements analysis	10	80%	/	/	30	67%	/	/	/	/	70%	7 (D)
International accounting	10	60%	/	/	30	37%	/	/	/	/	42,5%	5 (F)
Auditing	5	100%	15	80%	/	/	30	77%	4	63%	74%	8 (C)
Internal auditing	5	80%	15	73%	/	/	30	80%	3	64%	72%	8 (C)

“Q” refers to number of questions

“R %” refers to the response rate based on grade points (achieved points / max. total points)

“G” refers to grade (according to the official grading system for European Educational Institutions – ECTS grading scale)

(Source: Author’s calculation)

ChatGPT exhibited its highest performance in the subject of Principles of Accounting, followed closely by Auditing and Internal Auditing. However, it failed to attain a satisfactory passing score, falling below 50%, in subjects such as Management Accounting II, Governmental Accounting and Accounting of Non-Profit Organizations, and International Accounting.

Drawing preliminary conclusions from our research and in-depth analysis, several key observations emerge:

- ChatGPT, functioning as a language model, excels in providing well-structured responses to both short and extensive qualitative questions, but exhibits comparatively weaker performance when dealing with quantitative calculations.
- ChatGPT demonstrates remarkable proficiency in handling true/false questions and MCQs with a single correct answer.

- It performs better in multiple-choice questions (MCQs) with a single correct answer, as opposed to MCQs with multiple correct answers.

An intriguing observation is that ChatGPT's performance aligns with or mirrors student performance in some subjects, while showcasing starkly contrasting outcomes in others.

This study marks a pioneering endeavor in the realm of higher education in North Macedonia, representing the first of its kind in this context. The research outcomes are substantiated through empirical evidence and align with previous global research findings, as exemplified by Wood et al. (2023). Nevertheless, it serves as an inaugural exploration within a series of such studies. These preliminary findings establish a robust foundation for subsequent research endeavors, particularly in the context of individual subjects.

The research's overarching goal was to pinpoint initial reference points, which serve as the departure point for all forthcoming experiments. These points hold significant prominence due to the inherent dichotomy of AI tools—they pose substantial risks on one hand while simultaneously presenting significant opportunities and potential on the other. As such, they remain highly pertinent, current, and thought-provoking, underscoring the critical need for ongoing research and exploration in this dynamic field.

Limitations of the conducted experimental research warrant attention. Firstly, ChatGPT 3.5, utilized in the study, is a free tool, while research highlights the advancements in the latest version, ChatGPT 4 (as of September 2023), which offers enhanced capabilities. Additionally, ChatGPT's algorithm is based on information available only up to September 2021, while many curricula and exam questions in the analyzed subjects, such as accounting and auditing standards and interpretations, undergo continual updates. It's crucial to note that the evaluation and point allocation for open-ended short questions and essay questions (case studies) possess an inherent level of subjectivity, incorporating partial scoring.

For future research, we recommend a more focused exploration within specific areas. Furthermore, testing ChatGPT 4, which boasts documented empirical improvements, should be a priority. As educators, harnessing the latest AI tools to enhance the educational process is essential. Our recommendations align with those presented by Nunes et al. (2023) and Kolade et al. (2023), suggesting opportunities for educational improvement through participatory learning, personalized support, increased engagement, and thoughtful exam design that emphasizes logical thinking and comprehensive understanding over rote memorization and selective learning.

5. CONCLUSION

In the swiftly changing terrain of education and technology, the fusion of AI, with ChatGPT serving as a notable example, stands out as a potent catalyst for transformation. Our research, the first of its kind in higher education in North Macedonia, has ventured into uncharted territory to assess ChatGPT's efficacy in enhancing exam performance within the realm of accounting and auditing.

Through rigorous experimentation and analysis, we have uncovered both the potential and limitations of ChatGPT in the context of exam questions in higher education. It has demonstrated commendable capabilities in answering qualitative questions, true/false inquiries, and multiple-choice questions with single correct answer, effectively automating routine tasks and providing detailed explanations. However, it faces challenges in handling quantitative calculations and complex MCQs with multiple correct answers.

ChatGPT transcends its role as a tool exclusively designated for higher education, accounting, and auditing. Instead, it stands as a cutting-edge AI language model with versatile capabilities, apt for

an expansive array of natural language processing tasks, including question answering, text generation, explanation provision, and more. Consequently, due to its extensive knowledge base, ChatGPT offers robust explanations and responses to open-ended questions. In the domain of quantitative tasks, ChatGPT adeptly formulates and furnishes information on executing specific calculations. However, it falls short in delivering fully computed results.

Our findings align with broader discussions about AI's role in education worldwide. ChatGPT, as a free tool with certain constraints, serves as a valuable assistant rather than a complete replacement for traditional teaching and evaluation methods. It offers educators an opportunity to leverage technology to engage students more effectively and encourage logical thinking and deeper understanding.

ChatGPT represents a promising addition to the educational toolkit. While it is not without its limitations, its potential to augment the learning experience and assist both students and educators is undeniable. As the fields of AI and education continue to evolve, embracing these technologies thoughtfully can pave the way for more engaging and effective educational practices.

REFERENCES

- ACPM. (2023). Does Chat GPT threaten the accounting profession?, available at: <https://acpmglobal.co.uk/blog/item/46-chi-zagrozhuve-chat-gpt-profesi-yi-buhgaltera> (accessed at August 6, 2023).
- Al Ghatrifi, M. O. M., Al Amairi, S. S. J. & Thottoli, M. M. (2023). Surfing the technology wave: An international perspective on enhancing teaching and learning in accounting, *Computers and Education: Artificial Intelligence*, Vol. 4. <https://doi.org/10.1016/j.caeai.2023.100144>
- Beerbaum, O. D. (2023). Generative Artificial Intelligence (GAI) with Chat GPT for Accounting – a Business Case, *Special Issue on Generative Artificial Intelligence (GAI) and Accounting*. <http://dx.doi.org/10.2139/ssrn.4385651>
- Bommarito, J., Bommarito, J. M., Katz, J. & Katz, M. D. (2023). GPT as Knowledge Worker: A Zero-Shot Evaluation of (AI) CPA Capabilities, available at: <https://arxiv.org/pdf/2301.04408.pdf> (accessed August 15, 2023).
- Brownlee, M. (2023). What ChatGPT can do for accountants, *Accounting Times*, available at: <https://www.accountingtimes.com.au/technology/what-chat-gpt-can-do-for-accountants> (accessed at August 5, 2023).
- Constantz, J. (2023). Almost 30% of professionals say they've tried ChatGPT at work, available at: <https://www.bnnbloomberg.ca/almost-30-of-professionals-say-they-ve-tried-chatgpt-at-work-1.1872520> (accessed at August 24, 2023).
- DAIR.AI. (2023). Prompt Engineering Guide, available at: <https://www.promptingguide.ai/> (accessed at July 3, 2023).
- Eulerich, M., Sanatizadeh, A., Vakilzadeh, H. & Wood, D. A. (2023). Can Artificial Intelligence Pass Accounting Certification Exams? ChatGPT: CPA, CMA, CIA, and EA?, available at <http://dx.doi.org/10.2139/ssrn.4452175> (accessed august 25).
- Faggella, D. (2020). AI in the accounting big four - comparing Deloitte, PwC, KPMG, and EY" [WWW Document], *Emerj, The AI Research and Advisory Company*, available at: <https://emerj.com/ai-sector-overviews/ai-in-the-accounting-big-four-comparing-deloitte-pwc-kpmg-and-ey/> (accessed at July 26, 2023).
- Frey, C. B., and M. A. Osborne. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, Vol. 114, pp. 254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>

- George, S. A. & GEORGE, H. S. A. (2023). A Review of ChatGPT AI's Impact on Several Business Sectors, *Partners Universal International Innovation Journal*, Vol. 1, No. 1, pp. 9-23. <https://doi.org/10.5281/zenodo.7644359>
- Han, H., Shiwakoti, K. R., Jarvis, R., Mordi, C. & Botchie, D. (2023). Accounting and auditing with blockchain technology and artificial intelligence: A literature review, *International Journal of Accounting Information Systems*, Vol. 48. <https://doi.org/10.1016/j.accinf.2022.100598>
- Huq, U. S. (2014). The Role of Artificial Intelligence in the Development of Accounting Systems: A Review, *The IUP Journal of Accounting Research & Audit Practices*, Vol. 13, No. 2, pp. 7-19, <https://ssrn.com/abstract=2508591>
- Kolade, O., Owoseni, A. & Egbetokun, A. (2023). Assessment on trial? ChatGPT and the new frontiers of learning and assessment in higher education, available at: <https://doi.org/10.13140/RG.2.2.32665.49768> (retrieved August 17, 2023).
- Mathew, A. (2023). Is Artificial Intelligence a World Changer? A Case Study of OpenAI's Chat GPT. *Recent Progress in Science and Technology*, Vol. 5, pp. 35-42. <https://doi.org/10.9734/bpi/rpst/v5/18240D>
- Newton, M. P. & Xiromeriti, M. (2023). ChatGPT Performance on MCQ Exams in Higher Education. A Pragmatic Scoping Review, available at: <https://edarxiv.org/sytu3> (accessed August 20, 2023).
- Nunes, D., Primi, R., Pires, R., Lotufo, R. & Nogueira, R. (2023). Evaluating GPT-3.5 and GPT-4 Models on Brazilian University Admission Exams, available at: <https://arxiv.org/pdf/2303.17003.pdf> (accessed August 15, 2023).
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, *Internet of Things and Cyber-Physical Systems*, Vol. 3, pp. 121-154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Susnjak, T. (2022). CHATGPT: The end of online exam integrity?, available at: <https://arxiv.org/abs/2212.09292> (accessed at July 8, 2023).
- Tan, X. (2023). The Impact of ChatGPT on Education and Future Prospects, *Highlights in Science, Engineering and Technology*, Vol. 61, pp. 138-143. <https://doi.org/10.54097/hset.v61i.10285>
- Wood, A. D., et al. (2023). The ChatGPT Artificial Intelligence Chatbot: How Well Does It Answer Accounting Assessment Questions?, *Issues in Accounting Education*, Vol. 38, No. 4, pp. 1-28. <https://doi.org/10.2308/ISSUES-2023-013>
- Yan, X., Li, H. & Pratap, O. (2022). Informatization Teaching Reform of Accounting Courses Under Modern Information Technology, In: Sun, S., Hong, T., Yu, P., Zou, J. (eds) *Signal and Information Processing, Networking and Computers. ICSINC 2021. Lecture Notes in Electrical Engineering*, vol 895. Springer, Singapore. https://doi.org/10.1007/978-981-19-4775-9_84