Journal of Accounting Education 73 (2026) 101000

Contents lists available at ScienceDirect

Journal of Accounting Education

journal homepage: www.elsevier.com/locate/jaccedu

Can ChatGPT be a knowledgeable and accurate ethical accountant? Assessing the overall usefulness of ChatGPT's answers to ethical dilemmas☆

Fábio Albuquerque [a,b,*], Paula Gomes dos Santos [a,c]

[a] ISCAL/Instituto Politécnico de Lisboa, Lisboa, Portugal
[b] CICF, School of Management, IPCA, Barcelos, Portugal
[c] COMEGI, Universidades Lusíada, Lisboa, Portugal

ARTICLE INFO

Keywords:
Accounting
Accuracy
ChatGPT
Consistency
Ethical dilemmas
Reliability
Usefulness

ABSTRACT

This paper aims to assess the overall usefulness of ChatGPT regarding ethical dilemmas in accounting by exploring the tool's ability to produce accurate answers, as well as reliably and properly identify the relevant legal frameworks in its justification. This study employs a quasi-experimental method and an exploratory perspective to evaluate ChatGPT 4.0's responses to questions from the Portuguese Order of Certified Accountants exams. Consistency and robustness tests were conducted by varying and repeating the prompts, and the outputs (ChatGPT responses) were assessed through a content analysis method by comparing them with official answer keys (accuracy) and their justification (legal frameworks). The findings indicate that, throughout the process, ChatGPT's performance is influenced by prompt design and repetition, and it consistently lacks accuracy, namely when professional judgment is required. Besides, although it is generally capable of recognising the relevant legal frameworks, its justifications were usually vague. Moreover, whenever more precise answers were provided, evidence of hallucinations was commonly found. Therefore, while the responses are often persuasive and structured, their reliability is limited, raising concerns about the model's epistemic soundness and reducing its overall usefulness, which highlights the need for users to exercise caution when relying on their outputs. To the best of the authors' knowledge, this is the first research to analyse in depth the characteristics within ChatGPT responses, specifically regarding professional ethics in an accounting proficiency exam. This study offers empirical evidence on ChatGPT's limitations and potential as a supplementary tool in accounting ethics. It contributes to ongoing debates about the role of generative AI in professional education and practice, providing insights for educators and practitioners regarding its responsible and cautious integration. While promising as a discussion aid, ChatGPT still cannot replace the critical thinking and contextual reasoning required for ethical decision-making.

1. Introduction

Society's perception of the accounting profession's legitimacy has been hindered by several corporate misconduct cases (Carnegie & Napier, 2010), enhancing the importance of ethical business practices, particularly in the accounting field (Costa & Pinheiro, 2023; Xu, 2025). For instance, accountants are confronted with several ethical dilemmas to ensure the usefulness of reporting, which makes accounting one of the most ethically demanding professions (Todorović, 2018). For this reason, accounting professionals must have ethical values given that stakeholders rely on the information's trustworthiness to play its relevant role in society (Ariail et al., 2024; Payne et al., 2020). Thereby, the development of accounting ethics curricula in higher education and professional accounting institutions has been reinforced and addressed by the literature (e.g. Ariail et al., 2023; Cheung et al., 2023; Onumah & Owusu, 2024).

Concerning the profession, several accounting organisations developed professional conduct or ethics codes, such as the Association of International Certified Professional Accountants (AICPA) (AICPA, 2023) and the International Federation of Accountants (IFAC) through its International Ethics Standards Board for Accountants (IESBA) (IESBA, 2024). For instance, IFAC's professional ethics code sets out integrity,

☆ This article is part of a special issue entitled: 'Artificial Intelligence' published in Journal of Accounting Education.

* Corresponding author at: Avenida Miguel Bombarda, 20, 1069-035 Lisboa, Portugal.
E-mail addresses: fhalbuquerque@iscal.ipl.pt (F. Albuquerque), pasantos@iscal.ipl.pt (P.G. dos Santos).

https://doi.org/10.1016/j.jaccedu.2025.101000
Received 11 April 2025; Received in revised form 14 October 2025; Accepted 6 November 2025
Available online 9 November 2025

objectivity, professional competence and due care, confidentiality, and professional behaviour as fundamental principles of ethics for professional accountants (IESBA, 2024).

Notwithstanding whether professional conduct and ethics codes are important for guiding accountants through moral dilemmas, they have been seen as insufficient on their own since they require an application to specific situations and are based on the accountants' ethical principles (Ariail et al., 2023; Preuss, 1998). The lack of effectiveness of the professional ethics codes (Ariail et al., 2023) led some authors to propose ethical models for decision-making in the accounting profession, namely with a philosophical orientation (Dunn & Sainty, 2020; Payne et al., 2020).

Literature has highlighted that accountants' ethical judgments can be determined by several factors, such as age (Ariail et al., 2024; Marques & Azevedo-Pereira, 2009), gender (Ariail et al., 2012; Marques & Azevedo-Pereira, 2009), the seniority in the profession (Uyar et al., 2015), or social, economic-organisational and environmental factors (Namazi & Rajabdorri, 2020). For instance, older accountants are more concerned with moral values than younger ones (Ariail et al., 2024).

From a broader perspective, Stahl (2025) argues that moral judgments can be explained under the ethical theory as part of moral philosophy. In the context of the accounting academic curricula, this may represent a challenging aspect in higher institutions concerning how to incorporate ethics as a discipline. For instance, Highfiel (2020) stressed that the lack of philosophical ethics skills of the academics who taught ethics limits the students' interiorization of the professional values, ethics, and attitudes essential for accountants in a multicultural world. Those aspects gain more relevance given that the literature has additionally stressed that the (mis)alignment of personal values to professional values affects ethical behaviour (Ariail et al., 2020a). Moreover, the personal values of accounting students differ from those of the accounting profession (Ariail et al., 2020b; Ariail et al., 2021).

The accounting field, like many other industries, has been highly impacted by the development of Generative Artificial Intelligence (GenAI) technology, namely large language models (LLMs) (e.g., Al Naqbi et al., 2024; Boritz & Stratopoulos, 2023; Cheng et al., 2024; Eisikovits et al., 2024; Eulerich et al., 2024; Kayser & Telukdarie, 2024; Ross & Zhang, 2024; Ruan et al., 2024; Vasarhelyi et al., 2023; Zadorozhnyi et al., 2023; Zhao & Wang, 2024).

Regardless of the great potential of GenAI, the literature has been stressing several concerns regarding, among other aspects, the ethical risks associated with its use and the importance of maintaining human critical thinking (e.g., Al Naqbi et al., 2024; Dwivedi et al., 2023; Elbaz et al, 2024; Fachrurrozie et al., 2025; Matos et al., 2024; Naik et al., 2024; Ruan et al., 2024; Xu, 2025; Wach et al., 2023). Concerning ethical or moral judgments specifically, GenAI may even act as an advisor, but humans remain essential (Giubilini & Savulescu, 2018; Okamoto et al., 2025). A major limitation of GenAI models is that they rely on data quality and may produce biased outcomes while lacking diverse or representative training data, therefore limiting the deep contextual understanding needed for replacing human analytical thinking, namely in complex ethical or culturally sensitive discussions (Xu, 2025).

Currently, one of the most disseminated GenAI tools is ChatGPT (Al Naqbi et al., 2024), which is based on LLM technology. According to Wang et al. (2025), both LLMs' moral competence and human morality evolve according to Dewey's philosophy, i.e., from training. However, the first lacks self-reflection capabilities, preventing it from attaining the moral competence level of humans. Still according to those authors (Wang et al., 2025, p. 9), "ChatGPT lacks a deep understanding of complex ethical issues and their consequences, as it is not consistently capable of discerning harmful intentions in questions. Its generalizability in ethical practice is still limited, and it continues to face challenges in moral reasoning and logical consistency".

Thereby, although ChatGPT has the potential to learn and execute certain tasks outperforming humans (Retzlaff et al., 2024), literature has shown that it aids but cannot replace human intelligence (e.g., Abeysekera, 2024; Cohen et al., 2023; Okamoto et al., 2025; Stott & Stott, 2023; Tsai et al., 2024).

This concern has been highlighted in recent studies that have critically examined the effectiveness of GenAI in addressing ethical dilemmas in different fields. For example, studies assessing GenAI's performance using medical proficiency exams found that while ChatGPT shows potential as a decision-support tool in navigating ethical issues, it occasionally produced incorrect or misleading responses, made reasoning errors, or exhibited bias; therefore, its epistemic limitations warrant caution (Khan et al., 2025; Sareen, 2024). Moreover, although ChatGPT could identify and articulate key ethical concerns, it often struggled with the subtleties of ethical dilemmas and sometimes misapplied moral principles, thereby performing worse in terms of the depth and accuracy of the answers (Balas et al., 2024).

Regarding the usefulness of GenAI in addressing ethical issues within the accounting area, it has been a less-studied matter. The literature has primarily assessed its capability to respond to accounting proficiency exams covering a range of topics, including ethical matters (Albuquerque & Gomes dos Santos, 2024a; Freitas et al., 2024; Oliveira & Khatib, 2023; Pinto et al., 2025).

Moreover, findings on GenAI's capability to address ethical questions have still seemed contradictory. On one hand, based on the Brazilian Sufficiency Examination of the Federal Accounting Council, some studies have shown that ethics is one of the areas in which ChatGPT (versions 3, 3.5, and 4.0) performed best (Freitas et al., 2024; Oliveira & Khatib, 2023). However, Oliveira and Khatib (2023) found a 100 % accuracy but only studied four questions from one exam. Freitas et al. (2024) studied sixteen questions from four exams, with a medium accuracy of 81 %, ranging from 50 % to 100 %.

Pinto et al. (2025) assessed several LLMs performance concerning the Portuguese Order of Certified Accountants (OCC, in the Portuguese acronym) exams, also concluding that both ChatGPT and Gemini perform better on ethics matters than on management and financial accounting questions. Notwithstanding, the accuracy rates from the questions on ethics of the fifteen exams studied are 38 %, 53 % and 52 % from ChatGPT 3.5, ChatGPT 4.0, and Gemini, respectively.

On the other hand, Albuquerque and Gomes dos Santos (2024a) found that ChatGPT 4.0 did not prove to be more accurate than version 3.5 in responding to the OCC exams. Furthermore, they also evidenced that both versions effectively identified the main issues in the proposed topics and provided critical analysis and explanations of underlying principles. However, some responses were found to be inaccurate, particularly in ethical questions which require more professional judgment.

The previous findings show that the accuracy of GenAI in addressing ethical questions needs to be further researched. It should also be stressed that none of those studies, in Portugal or Brazil, have been specifically focused on ethical matters. Besides, they have usually tested the overall accuracy of those tools through a first and single attempt, with no further considerations regarding their global understanding of the subjects and the consistency of their answers, namely their performance throughout different attempts and the possible diversity of ways in which a question can be proposed.

Considering those gaps in the literature, this paper aims to assess how ChatGPT performs in specifically addressing accounting ethical dilemmas, conducting several consistency tests. To fulfil this purpose, the answers from ChatGPT 4.0 to a set of 99 questions inserted into the newest model of exams by OCC, exclusively dedicated to ethics, are used as a proxy for accounting ethical dilemmas. Then, it assesses, through a content analysis method, its overall usefulness by the accuracy of its answers and its ability to reliably and properly identify the relevant legal frameworks in its justification throughout the process.

The findings suggest that, at first sight, ChatGPT's responses globally understand the subjects proposed, providing applicable insights regarding the topics under discussion and the relevant legal frameworks

in its justifications. However, they were usually vague and, whenever more precise answers were provided, evidence of hallucinations was commonly found. Besides, the responses can usually be inaccurate and may be dependent on either the attempt or the approaches used to propose the questions, providing inconsistent results. Therefore, concerns about the accuracy of its answers and the reliability of its justification can limit its overall usefulness regarding its primary understanding, stressing the need for users to be cautious when using it.

Therefore, the findings of this study have important practical and social implications. In professional accounting contexts, ChatGPT may serve as a useful preliminary tool for exploring ethical dilemmas by identifying relevant themes and offering structured perspectives. However, its inconsistent reliability and accuracy and limited contextual judgment highlight the necessity of critical human oversight. This is particularly important in the accounting profession, where ethical decisions often involve complex and real-world variables that go beyond standardised rules or codes. From a social perspective, the study underscores the broader concern that the accessibility and persuasive nature of ChatGPT's responses may mislead less experienced users, reinforcing a superficial understanding of complex ethical issues. This is especially relevant in educational environments where GenAI integration should be carefully mediated by educators to promote critical reflection. Overall, while GenAI may support ethical awareness, it still cannot substitute for the professional judgment and moral insight developed through human experience.

This paper is structured in four sections. The next section identifies the research questions, materials and methods underlying the exploratory analysis proposed for this paper and is followed by the results section. Finally, the last section presents the discussion and conclusions, as well as the limitations and suggestions for future avenues.

The next section presents the research questions, as well as the material and methods underlying this study.

2. Research questions, material and methods

This study is based on a quasi-experimental method. The application of the OCC's professional ethics code corresponds to the object of analysis underlying this research. It uses an exploratory perspective and content analysis as a method to assess the accuracy level of ChatGPT's answers and whether its responses can properly identify the legal frameworks under assessment. By doing this, its overall usefulness is, therefore, being tested in this research. The analysis follows a process that submits this tool to several consistency tests.

To achieve this purpose, it uses a set of questions proposed by the OCC in its latest national exam of 26 October 2024 as an object, which are specifically related to professional ethics issues (ethical dilemmas) required by the OCC so that candidates may access the profession of certified accountant in Portugal under the first model of Regulation 363/2024 of 1 July 2024.

Since the advent of that Regulation, three different models have emerged for candidates to access the accounting profession in Portugal. According to the first model, the candidate must complete an internship as part of the course, attend training at the Order and take a final and mandatory exam on professional ethics issues. Then, the second model requires the candidate to complete a professional internship and a final exam on professional topics. Finally, based on the third model, professional experience is valued, and the candidate attends modular training at the OCC and undergoes a final assessment of the training modules.

The OCC is an IFAC member, and its professional ethics code is based upon the principles which underlie the International Code of Ethics for Professional Accountants issued by IESBA (OCC, 2023). Specifically, the OCC's professional ethics code is comprised of only eighteen articles throughout less than nine pages, as follows:

- Article 1: Scope of application;
- Article 2: General duties;
- Article 3: General professional ethical principles;
- Article 4: Independence and conflicts of interest;
- Article 5: Responsibilities;
- Article 6: Professional Competence;
- Article 7: Accounting principles and standards;
- Article 8: Relation with the Order and other entities;
- Article 9: Contract;
- Article 10: Confidentiality;
- Article 11: Information duties;
- Article 12: Rights regarding the entities for which the services are rendered;
- Article 13: Conflict of interest between entities for which the services are rendered;
- Article 14: Fees;
- Article 15: Return of documents;
- Article 16: Loyalty among certified accountants;
- Article 17: Professional ethical violation;
- Article 18: Professional societies of certified accountants, societies of certified accountants and multidisciplinary societies.

Therefore, it is also a principle-based professional ethics code. However, it is significantly shorter than the IESBA's one, even if the content specifically related to the auditors is excluded, considering that this profession is regulated in Portugal by a distinct Order (the Order of Statutory Auditors). Considering this, it is imperative to note that the OCC is continuously promoting several events, courses and information sessions aimed at disseminating educational content regarding how to properly apply, in practice, the certified accountant's professional ethics code, by regularly promoting several educational events.

Underlying the objective behind this study, the following research question (RQ) is proposed.

RQ: Can ChatGPT be a useful tool to analyse professional ethical dilemmas in accounting?

Furthermore, considering that the analysis of ethical dilemmas can be seen as a combination of the two-fold characteristics mentioned above that jointly contribute to reaching ChatGPT's usefulness, two further subquestions (SQ) are additionally proposed for a more in-depth exploration of the RQ, as follows:

SQ1: Can ChatGPT identify the correct answer (most complete and accurate) regarding professional ethical dilemmas in accounting?

SQ2: Can ChatGPT reliably and properly identify the relevant legal frameworks underlying the questions regarding professional ethical dilemmas in accounting?

Specifically, this research uses the questions regarding professional ethics proposed in the OCC exam as inputs. The exam covers several subjects included in the accountants' professional ethics code throughout the 25 questions proposed in each OCC's exam, with some also being related to topics covered by the national accounting and reporting standards. Four different versions of this exam were distributed, which are also available on the OCC's website (https://occ.pt/pt-pt/inscricao/arquivo/enunciados-de-exame-e-conteudos-programaticos). The questions are originally multiple-choice types, with four alternatives (A, B, C or D). Then, 100 questions would be assessed in this research (25 questions x 4 exams). However, since one of them was considered not valid by OCC and none of the alternatives was considered, a total of 99 different questions are available to perform the tests. The questions were inserted using the original language (Portuguese) of the exams, and ChatGPT's answers were equally obtained in the same language. Nonetheless, some of them will be freely translated into English in this paper for illustrative purposes.

Regarding **SQ1**, the OCC's and ChatGPT's answers are the output and source of comparative analysis. After inserting those questions (input) into that tool, the answers (output) provided by ChatGPT will be gathered and compared with those by OCC. The OCC's answers are defined as the control element, as it is likely the best proxy for what can be considered the valid and correct answer, also preventing any

subjectivity underlying this analysis. Version 4.0 of ChatGPT is used, as it is the latest freely available (albeit with some restrictions) version.

To assess **SQ2**, each ChatGPT answer was reviewed to determine whether it referred to the proper source in its responses, namely, the relevant frameworks underlying the question. The answers were initially divided into two groups, according to the contents of ChatGPT's justifications, as follows:

- first group (**1st group**): it includes the answers that provide conclusions through an overall presentation of the relevant framework, with no specific provisions regarding, for instance, the number of a given article that is necessary to understand its conclusion. The incorrectness rate under the **1st group** of answers may happen whenever ChatGPT possibly indicates a source that does not apply at all to the case (e.g., the indication of IFRS whenever it is not applicable instead of the Portuguese accounting and reporting standards).
- second group (**2nd group**): it includes the answers that provide conclusions through specific provisions regarding, for instance, the number of a given article that is necessary to understand its conclusion. The incorrectness rate for the answers within the **2nd group** was divided into two types, as follows:

  - first type (**1st type**): ChatGPT indicated a source that could be applied, but it was wrongly indicated (e.g., the indication of an incorrectly given article of a proper and applicable source);
  - second type (**2nd type**): ChatGPT indicated a correct source (e.g. the indication of a correct article of a proper source), but this was misunderstood by this tool.

Specifically, the **1st group** is focused on those cases for which ChatGPT provides a vaguer indication of legal frameworks used for its responses, while the assessment behind those **1st and 2nd types** under the **2nd group** only considered those cases where a specific article of a given legislation is mentioned (e.g., a specific article from the OCC's professional ethics code).

The objective behind this assessment was to distinguish the different levels of hallucinations or misunderstanding by ChatGPT, as follows: from those that are easily identified (incorrect legal frameworks from the **1st group**) to those that require more attention (incorrect legal frameworks from the **1st type** under the **2nd group**), and finally, from those that can only be found meticulously (incorrect legal frameworks from the **2nd type** under the **2nd group**).

This analysis will be performed regardless of whether the question is considered right or wrong under the SQ1 assessment, as it has a distinct objective. Furthermore, it excludes ChatGPT's overall interpretation of a given applicable legislation or unidentified or non-specified legal frameworks since this analysis would probably be highly subjective and redundant with the previous conclusions under SQ1. Nonetheless, cross-checking analysis will be performed to check whether the different patterns of answers, i.e., how ChatGPT may indicate the legal frameworks to provide its justifications, may be related or not to its accuracy.

Overall, the incorrectness rate from the SQ2 assessment may provide a complementary validation for the findings behind SQ1 by checking if ChatGPT can sometimes find the right answer by presenting, however, an improper justification. During this process, the authors independently reviewed and rated all responses, resolving any disagreements through consensus to ensure the reliability of the assessment.

The results for each SQ will be subject to several consistency performance tests, which comprise five rounds of input–output requests to ChatGPT, to check the robustness of the findings. Fig. 1 shows the process and the types and characteristics of questions over the five rounds.

Throughout the process of the SQs' assessments above, the original questions from the OCC exam were inserted into ChatGPT and compared during the first and second rounds. If neither did not correspond to the one considered correct by OCC, a third round is performed by asking ChatGPT the same question again for those cases. If the correct answer, according to the OCC proposition, is still not verified, a fourth round is performed by asking ChatGPT if that could be seen as equally valid and correct. Finally, whenever the correct answer is still not achieved, the OCC question is included as part of a statement that only needs to be validated by ChatGPT, asking the tool if this statement (with no alternative options proposed) can be considered as correct.

These rounds aim to assess the consistency and robustness of ChatGPT's performance under repeated prompts (namely, rounds 1 to 3), and by using a different prompt (a statement, and not a multiple-choice question). The prompt did not include any specific request regarding the level of the content or any specific element to be provided under its justification, since the objective was precisely to understand how this tool performs without this type of command, which is the object of analysis in this research through the SQ2 assessment. In a similar sense, no previous tests were performed since the objective behind this experiment was precisely to check its outputs when a regular user requested it to provide a similar answer, including its performance from one round to another. Besides, it is worth mentioning an inherent constraint behind this experiment, which relates to ChatGPT's previous knowledge of the authors' precedent requests, based on the users' accounts where those requests were made.

Appendix I to Appendix V provides a case (the second version of the exam, question 2) that illustrates the consistency process from the first to the fifth round, respectively. For this case, the correct answer provided by OCC is alternative d). In the first round, ChatGPT indicated c) as the correct answer. Then, option a) was indicated in the second round instead. Following, ChatGPT repeated option a) in the third round. Finally, this tool did not accept option d) in the fourth round and kept

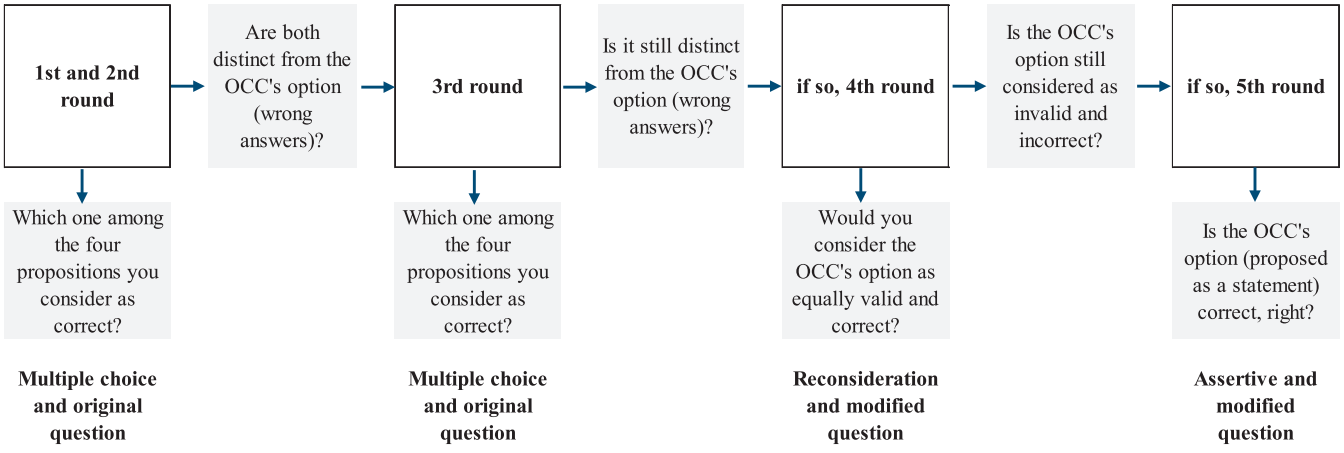

Fig. 1. Questions throughout the consistency tests, their types and characteristics.

refusing it as correct when the question was inserted as a statement to be validated in the fifth round.

The questions or cases (inputs) from the OCC exam were individually inserted (each question at a time) into ChatGPT on 18 May, with an underlying prompt that requires this tool to be focused on the Portuguese legal framework regarding professional ethics, as follows: "I would like to obtain answers to a set of questions that will be presented below and that include four possible alternatives, which must be answered following what is provided for in the OCC's statutes and professional ethics code of certified accountants in Portugal in force in 2024. I attach a document with its content. Among the possible alternatives, one and only one response option must be selected that, simultaneously, is most complete and best reflects the application of the aforementioned diplomas. The application of such diplomas must also be combined, whenever appropriate and pertinent, with the diverse legislation currently in force in Portugal in the different areas covered by the certified accountant's activities, namely in terms of tax, labour, social security, accounting and financial reporting and sustainability legislation, as well as in compliance with the civil, criminal, commercial and commercial company codes, public procurement, administrative procedures, among others".

The following answer was initially provided by ChatGPT: "Perfect! You can send me the questions you want me to analyse, along with the document with the OCC's statutes and the professional ethics code of certified accountants in force in 2024. This way, I will be able to:

- Evaluate each question individually;
- Identify the most complete and correct alternative based on the aforementioned diplomas and applicable Portuguese legislation;
- Justify, if you wish, your choice of the correct answer.

Please attach the document and submit the first question. I'm ready to start". After uploading the document, ChatGPT finally said the following: "Perfect, the document with the OCC's statutes and professional ethics code of certified accountants in force in 2024 is now available. You can now submit the questions you want me to review. For each one, I will indicate the correct alternative among the four presented, justifying the choice based on:

- In the OCC's statutes;
- In the OCC's professional ethics code;
- And, whenever necessary, in the complementary legislation applicable in Portugal (tax, labour, accounting, criminal, administrative, etc.).

I'll wait for the first question to start.".

This procedure is intended to give ChatGPT the same conditions that candidates have when they are submitted to this exam.

The next section provides the findings.

3. Results

This section presents the findings of each of the two SQs detailed in the previous one.

The findings for **SQ1** relate to the assessment of ChatGPT's accuracy by providing the correct (most complete and accurate) alternative, also taking into consideration the consistency tests (robustness tests) performed over this experiment. Although not significantly, ChatGPT provided the correct answers in 45 instances (or 45 %), while the same incorrect answers were verified 23 times (23 %). Besides, ChatGPT have changed its answers 31 times as a whole (31 % of cases). In addition, neither the size of the question nor the specific subject underlying it could be used to attribute a particular pattern to the cases for which the answers were coincident.

Therefore, the changes occurred in both directions, i.e., by changing a supposedly right answer to a wrong one or vice versa. Furthermore, ChatGPT also changed the wrong alternatives to another wrong one. Table 1 provides those findings.

Following, a third round was performed for those cases where wrong answers were found in both previous rounds (30 remaining cases, or 30 % of the initial cases). This confirmatory round solved nine cases out of 30 (30 %). This means that ChatGPT changed its initial answers again nine times, considering the starting point from the previous rounds, which were either different or the same, but equally wrong.

Table 1
Comparison of ChatGPT's answers from the first to the second round.

	Kept it right	Kept the same Wrong question	Right to Wrong	Wrong to Right	Wrong question to a different one	Total questions
Number	45	23	15	9	7	99
%	45	23	15	9	7	100

Therefore, the fourth round assessed the residual 21 cases by requesting this tool for a possible reconsideration to validate the OCC's answer. The pattern here is relatively clear. In contrast to the previous rounds, ChatGPT did not reconsider its initial position by doing these procedures. The answers were highly assertive, indicating this as impractical and providing justification for it.

There were only three cases in which this tool mentioned the OCC's choice as a possible correct answer but still inserted some cautions, which did not happen in its answers for the previous rounds. For instance, ChatGPT mentioned, for one of those questions, that "Option d) You have lost the right to civil liability insurance provided by the Order may seem plausible, but it is not sufficiently complete nor the most correct option in isolation, given the current regulatory framework of the Order of Certified Accountants (OCC). Let's analyse why: (…)".

Finally, whenever the questions were proposed as statements in the context of the fifth round, the results were inconsistent again, with no particular pattern to be stressed. More specifically, 11 cases out of 21 were changed from wrong to right (about 52 %).

Table 2 provides the findings throughout the rounds.

To sum up, it is relevant to consider that the best results obtained by ChatGPT during a single round were 61 % and 55 %, which occurred over the first and second rounds, respectively. Then, the questions proposed in an assertive tone during the fifth round achieved the third-best success rate (52 %). Conversely, the modified questions proposed during the fourth round, for ChatGPT to reconsider its initial answers, have shown the worst success rate, with no wrong previous responses being modified by this tool.

In turn, regarding **SQ2**, which intends to check whether ChatGPT can be useful by assessing the appropriate legal frameworks underlying the questions (inputs) and its reliability, the findings globally indicate its capability to identify it. Table 3 provides the correctness rate throughout the rounds.

Table 3 shows that, throughout the rounds, the first group of answers (based only on an overall framework) commonly covers the majority of the questions inserted, except for the third one, where the second group shares a similar number of records. The first and second round presents the highest percentages for those cases (70 % and 88 %, respectively). It was observed that, by introducing its proposed answers (outputs), ChatGPT typically mentions that the correct and best response is based on the Statutes and the professional ethics code for certified accountants in force in Portugal. Besides, whenever applicable, it commonly mentions further legislation and frameworks to be additionally considered according to the underlying issues, such as the tax, labour legislation, commercial codes, and accounting regulations in force in Portugal. This happens even when this legal framework is not directly included within the questions (input), i.e., it is only implicit in the ethical dilemmas proposed for ChatGPT's assessment. However, whenever it is not specifically requested through the prompt, it usually provides conclusions through an overall presentation of the relevant frameworks, being commonly vague, i.e., with no previous specifications of their respective applicable articles.

For instance, for a question related to the assets' depreciation, with

Table 2
Comparison of ChatGPT's answers throughout the rounds.

	Questions under assessment by round – Number	Right answers –Number (percentage)	Wrong answers –Number (percentage)
1st round	99	60 (61 %)	39 (39 %)
2nd round	99	54 (55 %)	45 (45 %)
3rd round	30	9 (30 %)	21 (70 %)
4th round	21	0 (0 %)	21 (100 %)
5th round	21	11 (52 %)	10 (48 %)

Table 3
Correctness rates of ChatGPT's answers in identifying the legal frameworks by groups and types of analysis, and rounds.

	First round	Second round	Third round	Fourth round	Fifth round
Questions under assessment by round	99 (100 %)	99 (100 %)	30 (100 %)	21 (100 %)	21 (100 %)
1st group – Number (In percentage)	69 (70 %)	87 (88 %)	15 (50 %)	13 (62 %)	12 (57 %)
Legal frameworks' correctness rate –Number (in % of first group)	66 (96 %)	87 (100 %)	15 (50 %)	13 (62 %)	11 (100 %)
Legal frameworks' incorrectness rate –Number (in % of first group)	3 (4 %)	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)
2nd group – Number (In percentage)	30 (30 %)	12 (12 %)	15 (50 %)	8 (38 %)	9 (43 %)
Legal frameworks' correctness rate –Number (in % of second group)	20 (67 %)	12 (100 %)	5 (33 %)	2 (25 %)	5 (56 %)
Legal frameworks' incorrectness rate –Number (in % of second group)	10 (33 %)	0 (0 %)	10 (67 %)	6 (75 %)	4 (44 %)
Of which:					
1st type incorrectness rate –Number (in % of second group)	7 (70 %)	0 (0 %)	4 (40 %)	3 (50 %)	3 (75 %)
2nd type incorrectness rate –Number (in % of second group)	3 (30 %)	0 (0 %)	6 (60 %)	3 (50 %)	1 (25 %)

Note: The **1st group** includes answers based on an overall framework without specific provisions. The **2nd group** includes answers based on specific provisions (e.g., article numbers). Within the **2nd group**, the **1st type** refers to wrongly indicated sources, and the **2nd type** to correctly indicated but misunderstood sources.

no reference to the national accounting standards, ChatGPT identified that "(…) According to the Accounting Standardization System (SNC, in the Portuguese Acronym) — particularly Accounting and Financial Reporting Standard (NCRF, in the Portuguese Acronym) 7 – Tangible Fixed Assets, the depreciation of an asset begins when the asset is available for use, that is, when it is in the necessary conditions to be used under the management's intention (…)".

It even describes specific contents from tax statements and other forms from several entities, such as the OCC and government authorities. For instance, a given answer mentioned that "(…) in the IES (Simplified Business Information) form, there is a specific field to indicate whether the accounts have already been approved at a general meeting or not. Therefore, it is technically admissible to submit the IES even before the formal approval of the accounts, if it is correctly marked in the appropriate field that the approval has not yet occurred (…)".

However, even these legal frameworks mentioned in a general way may be improperly presented. Table 3 shows that this happened in only 3 cases, which occurred during the first round. For instance, to a given question ChatGPT suggested that "(…) According to IAS 37 – Provisions, Contingent Liabilities and Contingent Assets, and equivalent national standards, a contingent liability must be disclosed in the explanatory notes when there is a possible, but not probable, possibility of an outflow of resources to settle a present obligation (…)". Despite not requiring a distinct application in these contexts, it is worthwhile to mention that IFRS can be used only for a reduced number of companies in Portugal or whenever the national accounting and financial standards have a significant gap to be solved, which was not the case for the example above and in the two other cases similarly found. Since those national standards can have differences when compared to the international ones issued by the International Accounting Standards Board (IASB), besides not proposing an adequate framework for analysis, the answers can be improperly provided.

Regarding the second group (answers based on specific provisions), as Table 3 also shows, the incorrectness rates are higher throughout the third, fourth and fifth rounds, where ChatGPT also increased the support in specific legislation. Except, again, the second round, where no incorrect answers were recorded, both types of proposed errors were found in this second group of answers, as follows: when ChatGPT indicated a source that could be used, but it was wrongly indicated (first type); and when ChatGPT indicated a correct source, but it was misunderstood by this tool (second type). This evidence indicates distinct levels of ChatGPT's hallucinations.

Appendix VI provides a comprehensive analysis of the issues identified from the content analysis of ChatGPT answers for those categories, by rounds.

As examples of the first type of incorrect legal frameworks, ChatGPT indicated that "(…) Article 64 of the OCC's statutes establishes that acts that disregard accounting and ethical standards constitute disciplinary infractions". However, the OCC's statutes regulate the certified accountants' application in Article 64, while what is mentioned by ChatGPT can only be found in Article 78. It must be stressed that both the OCC's statutes and professional ethics code were directly provided to (uploaded on) ChatGPT. Another example of a wrong reference occurred when ChatGPT mentioned that "(…) According to Article 3 of the Individual Income Tax (CIRS, in the Portuguese Acronym), income from dependent work includes all remuneration paid on account of an employment relationship, regardless of its regularity (…)". Instead, the CIRS describes the incomes from dependent work in Article 2. Conversely to the previous case, this was not a source directly provided to ChatGPT.

An example of the second type of errors is the mention of Article 10 of the Statutes as framing the duty of the certified accountant to "apply the correct technical bases and inform the company's management of irregularities or discrepancies", which is not expressly within the scope of the aforementioned article.

Curiously, the third and fourth rounds increased the percentage of the second type of incorrectness, which can possibly be explained by a ChatGPT strategy to convince that its response was right. For instance, it is justified that the VAT Code (article 78.7) establishes that: "Proof that the credits have become doubtful is provided through a declaration by the taxpayer accompanied by certification by a certified accountant or statutory auditor", when this claim does not exist in the VAT Code.

Furthermore, it is worth mentioning that all the contents are provided with an assertive tone and a reader-friendly presentation through the common use of bullets, which is a pattern that was found as a typical

characteristic of ChatGPT's answers throughout the process and regardless of the type of questions that were proposed (over the consistency tests). In each ChatGPT's response, only a unique option was considered by this tool, with no further alternatives to be reflected concerning a possible distinct context or interpretation. This happened even when the answers were changed by it from one round to another.

Moreover, no further information was requested by ChatGPT in any case, which means that this tool never considered a possible case of missing information or misunderstanding. For instance, when responding to the multiple options question, ChatGPT begins a given answer by indicating that "The correct alternative is b) You can invoke the just impediment and proceed with sending the declaration until July 29th (…)", then providing its justification. At the end, it also indicates why the further options should not be considered. For instance, and considering the same question, ChatGPT responded that they were incorrect by considering the following reasons: "(…) a) Incorrect — The just impediment can be invoked even if it occurs on the same day as the obligation, according to the OCC's statutes; c) Incorrect — July 25 would be just 10 exact days, but does not include weekend tolerance, nor does it correspond to the maximum period allowed; d) Incorrect — August 31 would only be applicable in cases of prolonged impediment (art. 12-B), which does not occur in this scenario (…)". Curiously, at the third round, ChatGPT changed its answer to c), although maintaining the same explanation saying that "Since the original deadline for submitting the IES ends on July 15th, the new deadline (adding 10 business days) ends on July 25th".

On the other hand, it usually suggests a further exploration of the matters under assessment at the end of each answer, as an invitation to users to know more about the subjects (topics and sources) or even to practitioners to solve some practical aspects. For instance, for the same question above, ChatGPT recommends the following: "(…) If you wish, I can help you write the declaration of just impediment or check the correct form on the Tax Portal. Do you want to continue with question 9?".

Moreover, some interesting characteristics underlying the ChatGPT responses were observed during the process of SQ2 assessment. Firstly, there is no evident pattern to be highlighted concerning the size of the ChatGPT answers. Therefore, the size of the questions does not necessarily result in wordier responses from ChatGPT. Following, specific responses (with articles of a given legislation) usually derive from specific questions (for instance, questions that also mention the articles). However, the opposite is not necessarily true, i.e., specific questions do not necessarily lead to specific answers. Thirdly, the pattern of ChatGPT responses, even for the same questions, may change throughout the rounds, i.e., it is inconsistent, regarding either the structure or the size. Then, the structure and size of ChatGPT responses also changed to the extent that the inputs increased (the number of sequential consults over a given round), being tendentially shorter and non-specific.

Finally, it should be noted that the incorrect legal framework is given by ChatGPT regardless of the answers being right or wrong. For instance, in the first round, 6 out of 10 answers with incorrect legal frameworks of the second group have supported its right answers.

The last section discusses the findings and provides the conclusions from this paper, including its limitations and suggestions for future research.

4. Discussion and conclusion

Using a quasi-experimental methodology and an exploratory perspective, this study aimed to assess ChatGPT's usefulness in addressing ethical dilemmas in accounting (RQ), by examining whether ChatGPT is capable of (i) producing accurate responses to such dilemmas (SQ1) and (ii) identifying the appropriate legal frameworks underpinning ethical questions (SQ2). The usefulness of the answers provided by ChatGPT was subject to consistency tests (robustness tests), by repeating the questions and by proposing them in different ways. To achieve this, responses from ChatGPT 4.0 to a set of 99 questions from the latest ethics-focused examination model by the OCC are used as a proxy for accounting ethical dilemmas. The inputs (questions) in ChatGPT were extracted from the OCC exam, and the outputs (answers) were evaluated through a content analysis method by comparing them to the proposed key for the SQ1 assessment, while the content of ChatGPT answers was the object of analysis for the SQ2.

The findings for **SQ1** show that, although ChatGPT produced persuasive, comprehensive and structured responses, its accuracy was globally reduced and inconsistent, particularly when questions required higher levels of professional judgment or technical interpretation. These limitations are consistent with the concerns highlighted in prior literature, which notes that GenAI tools lack the contextual depth and self-reflective reasoning required for ethical analysis (Giubilini & Savulescu, 2018; Xu, 2025; Wang et al., 2025). As observed, the variance in accuracy raises questions about ChatGPT's usefulness as a whole, which is dependent not only on the many times questions were posed, but also on how they were posed. Specifically, no particular pattern was found that, once known, could be mitigated, except in those cases where ChatGPT seemed not to reconsider its previous answers throughout confirmatory questions (over the fourth round). The evidence of such constraint, proposed in this research through its several consistency tests, may likely explain the variability (different findings) in proficiency exam studies found in the literature (Freitas et al., 2024; Pinto et al., 2025), where accuracy rates in ethics-related questions ranged widely across different GenAI models and versions.

Regarding **SQ2**, the findings suggest that ChatGPT generally succeeds in identifying the appropriate legal frameworks underlying ethical accounting questions. This is aligned with previous studies (e.g., Albuquerque & Gomes dos Santos, 2024a), which found that ChatGPT was capable of recognising key ethical issues and articulating relevant frameworks. Similar to insights from other domains (Khan et al., 2025; Balas et al., 2024), ChatGPT demonstrated an ability to provide critical analysis and explanation of legal frameworks. Nonetheless, it also tends to provide vague justifications in its answer whenever the prompt does not specifically call for a higher-level specification. Besides, while the model can frame ethical content, its performance is not entirely reliable, particularly regarding the specific articles indicated and dubious interpretations. Therefore, while ChatGPT often cited relevant Portuguese regulations and frameworks, it misattributed articles or relied, sometimes, on inappropriate standards such as IFRS, which are only conditionally relevant in the Portuguese context. Thus, this finding suggests that even when ChatGPT appears well-informed, users must critically assess its references and contents, considering the evidence of different levels of ChatGPT's hallucinations. Also, the incorrect legal framework is provided by ChatGPT regardless of whether the answers are right or wrong. Therefore, decisions based on its use must be cautious, which contrasts with its high assertiveness, which is particularly unsuitable for users with a low level of knowledge in a given matter.

Moreover, it should be noted that ChatGPT's responses are mediated by prompt design and repetition. Whether by repeating the prompt or by changing it, namely by making a statement instead of a question, different responses were obtained. However, even when making a statement, although some responses were altered to the right ones (maybe due to the tool's tendency to conform to users' desired outcomes), some of them remain wrong, showing ChatGPT's resistance to correction, which may reflect a form of "epistemic stubbornness". Even after repeated prompts and provision of the correct answer, the model frequently failed to adjust its position or acknowledge alternative correct interpretations. This rigidity, combined with its hallucinations and highly persuasive tone, underscores the key risk of ChatGPT sounding confident even when it is demonstrably wrong. This suggests that the model's usefulness is limited not only by epistemic gaps but also by an inability to engage in reflective revision, especially in the absence of strong corrective signals in the prompt. This rigidity is especially concerning in ethical reasoning, which often requires context sensitivity.

Therefore, although ChatGPT can serve as a valuable starting point for ethical accounting analysis, its usefulness is diminished by its low level of epistemic soundness and accuracy, both of which converge to what can be seen as a compromised reliability. These constraints are particularly critical in the accounting profession, where ethical decisions are shaped by an interplay of personal, organisational, and environmental factors (Ariail et al., 2024; Ariail et al., 2012; Marques & Azevedo-Pereira, 2009; Namazi & Rajabdorri, 2020; Uyar et al., 2015). Unlike human professionals, whose ethical reasoning develops through experience and self-awareness (Wang et al., 2025), ChatGPT's responses depend heavily on training data and prompt formulation, often lacking the nuance required to resolve moral ambiguity.

Moreover, the findings reiterate the broader concern in accounting ethics literature that ethical standards, though essential, are not sufficient. As Ariail et al. (2023) and Preuss (1998) suggest, ethical codes must be interpreted and applied in context, relying on human judgment. The present study confirms that ChatGPT is not a substitute for such judgment. Instead, it may serve as a complementary tool that can assist in identifying relevant topics and offering structured perspectives, but whose outputs require critical scrutiny by knowledgeable professionals.

In conclusion, ChatGPT shows potential as a supplementary aid in analysing professional ethical dilemmas in accounting, particularly in recognising key issues and articulating general principles. These findings should also be considered in light of the current context of the initial stage of this tool development, which has been subject to rapid improvements that will likely mitigate these constraints in the coming years. Nonetheless, due to its variable accuracy and dependency on prompt design, it should currently be used with caution, especially in contexts where ethical decisions carry significant professional or societal implications.

The experimental analysis underlying this paper may contribute to academia and practitioners through different perspectives. Firstly, the findings from this paper can help accounting professionals understand that ChatGPT might serve as a brainstorming tool or starting point for discussing ethical dilemmas, but should not be relied upon for definitive guidance. Additionally, it can be integrated into classrooms to foster open discussions, encouraging critical thinking and source verification under the careful guidance of educators, making it a valuable resource for academic purposes and enhancing the learning process. In both cases, however, reliance must be tempered by awareness of its limitations.

Therefore, significant constraints for its use remain and must be considered by academics and practitioners since ChatGPT, despite may being efficient and useful, still cannot replace human perspective concerning either critical judgment or subjective reasoning.

This research has inherent limitations. These constraints concern the dependency on the information provided by users regarding the prompts, previous tests and possible adjustments proposed, the contents underlying the questions (inputs), and even how and how many times they are inserted, which always have a subjective or unsurpassable nature in such studies. Specifically, it must be considered that ChatGPT can generate different responses to the same question according to those elements, which researchers cannot totally control, and, therefore, may distinctly influence the findings. Besides, ChatGPT's previous knowledge of the users' precedent requests, based on its account registers, can be seen as a relevant constraint to be taken into account. Finally, the analysis of the findings was based on a by round-basis analysis. Nonetheless, it is worth mentioning that some relevant differences regarding the number of questions assessed by round can mitigate the validity of some comparisons made.

As future avenues, researchers can replicate this experimental study on ethical issues by using other alternative GenAI tools, such as Gemini, as well as in different countries, namely, to check the potential influence of other legal environments. The analysis of the students' and practitioners' reactions or acceptance of ChatGPT's answers could also be of interest. An experiment of such nature could lead researchers to identify how convincing those AI tools can be according to the type of users and their demographic characteristics, which can provide evidence of how useful they can be as a first source of information to be integrated into their critical analysis and reflections or, from a different perspective, how risky their applied use can be.

AI use Disclosure statement

Artificial intelligence tools were used during the preparation of this manuscript only to perform the tests, as it is provided in the methodological section. The final content was critically reviewed, edited, and approved by the authors, who take full responsibility for the integrity and accuracy of the work. No AI tools were used to generate original research data or to make analytical decisions.

CRediT authorship contribution statement

Fábio Albuquerque: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Paula Gomes dos Santos:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Funding

This work was supported by the Instituto Politécnico de Lisboa, project IPL/IDI&CA2024/IAccount_ISCAL [Grant number is not applicable].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to than the Instituto Politécnico de Lisboa for supporting this research. This study was conducted at the Research Center on Accounting and Taxation (CICF) and was funded by the Portuguese Foundation for Science and Technology (FCT) through national funds with the reference UID/04043/2025 and doi https://doi.org/10.54499/UID/04043/2025.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jaccedu.2025.101000.

Data availability

All data for this study are publicly available. Notwithstanding, the authors are available to share the data collected for this purpose upon reasonable request, namely for research purposes. If you are interested in this data, please contact the corresponding author.

References

Abeysekera, I. (2024). ChatGPT and academia on accounting assessments. *Journal of Open Innovation: Technology, Market, and Complexity, 10*(1). https://doi.org/10.1016/j.joitmc.2024.100213

AICPA. (2023). AICAP Code of Professional Conduct, Effective December 15, 2014. Updated for all Official Releases through December 2023. Available at https://pub.aicpa.org/codeofconduct/Ethics.aspx?vngagetrans=qzhZcZPAuJgPRyCdB9wY.

Al Naqbi, H., Bahroun, Z., & Ahmed, V. (2024). Enhancing Work Productivity through Generative Artificial Intelligence: A Comprehensive Literature Review. *Sustainability, 16*, 1166. https://doi.org/10.3390/su16031166

Albuquerque, F., & Gomes dos Santos, P. (2024a). Can ChatGPT Be a Certified Accountant? Assessing the Responses of ChatGPT for the Professional Access Exam in Portugal. *Administrative Sciences, 14*(7), 152. https://doi.org/10.3390/admsci14070152

Albuquerque, F., & Gomes dos Santos, P. (2024b). Exploring ChatGPT's capabilities in solving accounting standards problems: The case of IAS 37. *Cogent. Education, 11*(1). https://doi.org/10.1080/2331186X.2024.2412492

Ariail, D. L., Smith, K. T., Smith, L. M., & Khayati, A. (2024). An Examination of Ethical Values of Management Accountants. *Journal of Business Ethics, 195*(2), 407–423. https://doi.org/10.1007/s10551-024-05640-z

Ariail, D.L., Abdolmohammadi, M.J. & Murphy Smith, L. (2012). Ethical Predisposition of Certified Public Accountants: A Study of Gender Differences, Jeffrey, C. (Ed.) *Research on Professional Responsibility and Ethics in Accounting,* 16, 29-56. Doi: 10.1108/S1574-0765(2012)0000016005.

Ariail, D. L., Smith, K. T., & Smith, L. M. (2021). A pedagogy for inculcating professional values in accounting students: Results from an experimental intervention. *Issues in Accounting Education, 36*(4), 5–40. https://doi.org/10.2308/ISSUES-19-018

Ariail, D. L., Smith, K. T., Smith, L. M., Steyn, R., & Khayati, A. (2023). Improving ethical compliance in accounting and business: An ethics code focused value self-confrontation approach. *Research on Professional Responsibility and Ethics in Accounting, 26*, 23–54. https://doi.org/10.1108/S1574-076520240000026002

Ariail, D. L., Smith, K., & Smith, L. M. (2020). Do United States accountants' personal values match the profession's values (ethics code)? *Accounting, Auditing & Accountability Journal, 35*(5), 1047–1075. https://doi.org/10.1108/AAAJ-11-2018-3749

Ariail, D.L., Smith, K.T. and Smith, L.M. (2020b). Do American Accounting Students Possess the Values Needed to Practice Accounting?, Baker, C.R. (Ed.) *Research on Professional Responsibility and Ethics in Accounting* , 23, 63-89. https://doi.org/10.1108/S1574-.

Balas, M., Wadden, J. J., Hébert, P. C., Mathison, E., Warren, M. D., Seavilleklein, V., Wyzynski, D., Callahan, A., Crawford, S. A., Arjmand, P., & Ing, E. B. (2024). Exploring the potential utility of AI large language models for medical ethics: An expert panel evaluation of GPT-4. *Journal of medical ethics, 50*(2), 90–96. https://doi.org/10.1136/jme-2023-109549

Boritz, J. E., & Stratopoulos, T. C. (2023). AI and the Accounting Profession: Views from Industry and Academia. *Journal of Information Systems, 37*(3). https://doi.org/10.2308/ISYS-2023-054

Carnegie, G. D., & Napier, C. J. (2010). Traditional accountants and business professionals: Portraying the accounting profession after Enron. *Accounting, Organizations and Society, 35*(3), 360–376. https://doi.org/10.1016/j.aos.2009.09.002

Cheng, X., Dunn, R., Holt, T., Inger, K., Jenkins, J. G., Jones, J., Long, J. H., Loraas, T., Mathis, M., Stanley, J., & Wood, D. A. (2024). Artificial intelligence's capabilities, limitations, and impact on accounting education: Investigating ChatGPT's performance on educational accounting cases. *Issues in Accounting Education, 39*(2), 23–47. https://doi.org/10.2308/ISSUES-2023-032

Cheung, R.-W.-Y., Agrawal, R. K., & Choudhry, S. (2023). Do virtues matter? Accounting ethics education in Hong Kong. *International Journal of Ethics and Systems, 39*(4), 679–696. https://doi.org/10.1108/IJOES-11-2021-0207

Cohen, S., Manes Rossi, F., & Brusca, I. (2023). Debate: Public sector accounting education and artificial intelligence. *Public Money & Management, 43*, 725–776. https://doi.org/10.1080/09540962.2023.2209290

Costa, A.J., & Pinheiro, M.M. (2023). The Moral Competence of Portuguese Accounting Students. *Iberian Conference on Information Systems and Technologies*, CISTI, 2023-June. https://doi.org/10.23919/CISTI58278.2023.10211707.

Dunn, P., & Sainty, B. (2020). Professionalism in accounting: A five-factor model of ethical decision-making. *Social Responsibility Journal, 16*(2), 255–269. https://doi.org/10.1108/SRJ-11-2017-0240

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., & Wright, R. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eisikovits, N., Johnson, W.C., & Markelevich, A. (2024). Should Accountants Be Afraid of AI? Risks and Opportunities of Incorporating Artificial Intelligence into Accounting and Auditing. *Accounting Horizons.* https://doi.org/10.2308/HORIZONS-2023-042.

Elbaz, A. M., Salem, I. E., Darwish, A., Alkathiri, N. A., Mathew, V., & Al-Kaaf, H. A. (2024). Getting to know ChatGPT: How business students feel, what they think about personal morality, and how their academic outcomes affect Oman's higher education. *Computers and Education: Artificial Intelligence, 7*, Article 100324. https://doi.org/10.1016/j.caeai.2024.100324

Eulerich, M., Sanatizadeh, A., Vakilzadeh, H., & Wood, D. A. (2024). Is it all hype? ChatGPT's performance and disruptive potential in the accounting and auditing industries. *Review of Accounting Studies, 29*, 2318–2349. https://doi.org/10.1007/s11142-024-09833-9

Fachrurrozie, F., Nurkhin, A., Santoso, J. T. B., Mukhibad, H., & Wolor, C. W. (2025). Exploring the use of artificial intelligence in Indonesian accounting classes. *Cogent Education, 12*(1), Article 2448053. https://doi.org/10.1080/2331186X.2024.2448053

Freitas, M. M., Sallaberry, J. D., Silva, T. B. J., & Rosa, F. S. (2024). Application of ChatGPT 4.0 for Solving Accounting Problems. *The Journal of Globalization, Competitiveness, and Governability,* 18, 49–64. https://doi.org/10.58416/GCG.2024.V18.N2.03.

Giubilini, A., & Savulescu, J. (2018). The Artificial Moral Advisor. The "Ideal Observer" Meets Artificial Intelligence. *Philos. Technol., 31*, 169–188. https://doi.org/10.1007/s13347-017-0285-z

Highfiel, C. (2020). Ethics Education for Aspiring Professional Accountants. *Research in Ethical Issues in Organizations, 24*, 177–183. https://doi.org/10.1108/S1529-209620200000024016

IESBA (2024). Handbook of the International Code of Ethics for Professional Accountants. 2024 Edition. https://ifacweb.blob.core.windows.net/publicfiles/2025-02/2024%20IESBA%20HB_ENG_August%202024_Final.pdf.

Kayser, K., & Telukdarie, A. (2024). Literature Review: Artificial Intelligence Adoption Within the Accounting Profession Applying the Technology Acceptance Model (3). *Springer Proceedings in Business and Economics, 217–231*. https://doi.org/10.1007/978-3-031-46177-4_12

Khan, A.A., Khan, A.R., Munshi, S., Dandapani, H., Jimale, M., Bogni, F.M., & Khawaja, H. (2025). Assessing the performance of ChatGPT in medical ethical decision-making: a comparative study with USMLE-based scenarios. *Journal of Medical Ethics*, jme-2024-110240. https://doi.org/10.1136/jme-2024-110240.

Marques, P. A., & Azevedo-Pereira, J. (2009). Ethical Ideology and Ethical Judgments in the Portuguese Accounting Profession. *J Bus Ethics, 86*, 227–242. https://doi.org/10.1007/s10551-008-9845-6

Matos, E. J., Bertoncini, A. L. C., Figueiredo, M. C., Ames, D. C., & Serafim, M. C. (2024). The (lack of) ethics at generative AI in Business Management education and research [A (falta da) ética na inteligência artificial generativa no ensino e pesquisa em Administração]. *Revista de Administração Mackenzie, 25*(6), Article eRAMD240061. https://doi.org/10.1590/1678-6971/eRAMD240061

Naik, D., Naik, I., & Naik, N. (2024). Leveraging the Use of ChatGPT: Exploring Its Real-World Applications Including Their Related Ethical and Regulatory Considerations. In: Naik, N., Jenkins, P., Prajapat, S., Grace, P. (eds) Contributions Presented at The International Conference on Computing, Communication, Cybersecurity and AI, July 3–4, 2024, London, UK. C3AI 2024. Lecture Notes in Networks and Systems, vol 884. Springer, Cham. https://doi.org/10.1007/978-3-031-74443-3_38.

Namazi, M., & Rajabdorri, H. (2020). A mixed content analysis model of ethics in the accounting profession. *Meditari Accountancy Research, 28*(1), 117–138. https://doi.org/10.1108/MEDAR-07-2018-0365

OCC. (2023). Statutes and Professional Ethics Code of Certified Accountants [Estatuto da Ordem dos Contabilistas Certificados e Código Deontológico dos Contabilistas Certificados]. Available at https://www.occ.pt/fotos/editor2/estatutos2023_dig.pdf.

Okamoto, S., Kataoka, M., Itano, M., & Sawai, T. (2025). AI-based medical ethics education: Examining the potential of large language models as a tool for virtue cultivation. *BMC Medical Education, 25*(1), 185. https://doi.org/10.1186/s12909-025-06801-y

Oliveira, J., & Khatib, A. (2023). Man or Machine? An Exploratory Study of the Performance of Chat GPT 3.5 in the CFC Sufficiency Exam. Available at SSRN: https://ssrn.com/abstract=4560434 or https://doi.org/10.2139/ssrn.4560434.

Onumah, R. M., & Owusu, G. M. Y. (2024). What drives the attainment of goals of ethical education in higher institutions? The perception of professional accountants and accounting educators". *Journal of Applied Research in Higher Education, 16*(5), 1548–1563. https://doi.org/10.1108/JARHE-02-2023-0080

Payne, D. M., Corey, C., Raiborn, C., & Zingoni, M. (2020). An applied code of ethics model for decision-making in the accounting profession. *Management Research Review, 43*(9), 1117–1134. https://doi.org/10.1108/MRR-10-2018-0380

Pinto, A. S., Abreu, A., Costa, E., & Paiva, J. (2025). AI in Accounting: Can AI Models Like ChatGPT and Gemini Successfully Pass the Portuguese Chartered Accountant Exam? *Lecture Notes in Networks and Systems. LNNS, 859*, 429–438. https://doi.org/10.1007/978-3-031-78155-1_40

Preuss, L. (1998). On ethical theory in auditing. *Managerial Auditing Journal, 13*(9), 500–508. https://doi.org/10.1108/02686909810245910

Retzlaff, C. O., Das, S., Wayllace, C., Mousavi, P., Afshari, M., Yang, T., Saranti, A., Angerschmid, A., Taylor, M. E., & Holzinger, A. (2024). Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements. *Challenges, and Opportunities, Journal of Artificial Intelligence Research, 79*, 359–415. https://doi.org/10.1613/jair.1.15348

Ross, M., & Zhang, J. (2024). ChatGPT is Ready for the Profession—But is the Profession Ready for ChatGPT? *Accounting Horizons*. https://doi.org/10.2308/HORIZONS-2023-087

Ruan, R.-B., Zhu, Z.-P., Chen, W., & Li, W.-P. (2024). The relationship between ethical governance of government in science and technology and science and technology for social good of AI enterprises. *Studies in Science of Science*, 42(8), 1577-1586 and 1606. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85212339841&partnerID=40&md5=cd7918d062e2adf1052ce3d36669db7f.

Sareen, K. (2024). Assessing the ethical capabilities of Chat GPT in healthcare: A study on its proficiency in situational judgement test. *Innovations in Education and Teaching International, 61*(6), 1330–1340. https://doi.org/10.1080/14703297.2023.2258114

Stahl, B. C. (2025). Locating the Ethics of ChatGPT—Ethical Issues as Affordances in AI Ecosystems. *Information, 16*, 104. https://doi.org/10.3390/info16020104

Stott, F.A., & Stott, D.M. (2023). A Perspective on the Use of ChatGPT in Tax Education, Calderon, T.G. (Ed.) *Advances in Accounting Education: Teaching and Curriculum Innovations* (*Advances in Accounting Education, 27,* Emerald Publishing Limited, Leeds, 145-153. 10.1108/S1085-462220230000027007.

Todorović, Z. (2018). Application of Ethics in the Accounting Profession with an Overview of the Banking Sector, *Journal of Central Banking Theory and Practice*, ISSN

2336-9205. *De Gruyter Open, Warsaw, 7*(3), 139–158. https://doi.org/10.2478/jcbtp-2018-0027

Tsai, C.-Y., Hsieh, S.-J., Huang, H.-H., Deng, J.-H., Huang, Y.-Y., & Cheng, P.-Y. (2024). Performance of ChatGPT on the Taiwan urology board examination: Insights into current strengths and shortcomings. *World Journal of Urology, 42*(1), 250. https://doi.org/10.1007/s00345-024-04957-8

Uyar, A., Kuzey, C., Güngörmüs, A. H., & Alas, R. (2015). Influence of theory, seniority, and religiosity on the ethical awareness of accountants. *Social Responsibility Journal, 11*(3), 590–604. https://doi.org/10.1108/SRJ-06-2014-0073

Vasarhelyi, M. A., Moffitt, K. C., Stewart, T., & Sunderland, D. (2023). Large Language Models: An Emerging Technology. *Accounting. Journal of Emerging Technologies in Accounting, 20*(2), 1–10. https://doi.org/10.2308/JETA-2023-047

Wach, K., Duong, C.D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkiewicz, J., & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2), 7 - 30. https://doi.org/10.15678/EBER.2023.110201.

Wang, G., Wang, W., Cao, Y., Teng, Y., Guo, Q., Wang, H., Lin, J., Ma, J., Liu, J., & Wang, Y. (2025). Possibilities and challenges in the moral growth of large language models: A philosophical perspective. *Ethics Inf Technol, 27*, 9. https://doi.org/10.1007/s10676-024-09818-x

Xu, X. (2025). Comparative Analysis of GPT-4o and GPT-4.0 in Business Ethics Role-Play Simulations. In *ICAITE 2024–2024 International Conference on Artificial Intelligence and Teacher Education* (pp. 57–63). https://doi.org/10.1145/3702386.3702388

Zadorozhnyi, Z.M., Muravskyi, V., Pochynok, N., Muravskyi, V., Shevchuk A., & Majda M. (2023). Application of Chatbots with Artificial Intelligence in Accounting. *Proceedings - International Conference on Advanced Computer Information Technologies, ACIT*, 196–200. https://doi.org/10.1109/ACIT58437.2023.10275395.

Zhao, J., & Wang, X. (2024). Unleashing efficiency and insights: Exploring the potential applications and challenges of ChatGPT in accounting. *Journal of Corporate Accounting and Finance, 35*(1), 269–276. https://doi.org/10.1002/jcaf.22663