Ilomata International Journal of Tax and Accounting

Check for updates

OPEN ACCESS

RECEIVED 24 April 2026
ACCEPTED 16 July 2026
PUBLISHED 31 July 2026

CITATION
Supriyadi, Wardana AB, Kusuma IGKCBA, Nasikhudin, (2026). Artificial Intelligence-Assisted Financial Statement Analysis and Tax Risk Assessment: Evidence from a Quasi-Experimental Study. Ilomata International Journal of Tax and Accounting. 7 (3), 1-8.
doi: 10.61194/ijtc.v7i3.2374

TYPE Original Research
PUBLISHED 31 July 2026
DOI 10.61194/ijtc.v7i3.2374
VOL 7 Issue 3 July 2026

Artificial Intelligence-Assisted Financial Statement Analysis and Tax Risk Assessment: Evidence from a Quasi-Experimental Study

Supriyadi[1], Arief Budi Wardana[2], I Gede Komang Chahya Bayu Anta Kusuma[3], Nasikhudin[4]

[123]Polytechnic of State Finance STAN, Banten, Indonesia, [4]Directorate General of Taxes, Jakarta, Indonesia

Correspondence: priesoekarno@pknstan.ac.id[1]

Abstract

The increasing complexity of accounting–tax differences and risk-based tax administration has strengthened the need for more structured approaches to tax risk assessment. Financial statement analysis is widely used to identify potential tax risk signals; however, manual interpretation often produces inconsistent outcomes because financial indicators are highly interconnected and difficult to evaluate systematically. This study examines whether financial statement indicators can identify potential tax risk and whether Artificial Intelligence-assisted analysis improves the quality of tax risk identification compared with manual analysis. Using a quasi-experimental design with a difference-in-differences approach, this study evaluated the analytical performance of 130 Diploma III Tax students analyzing the financial statements of 25 Indonesian publicly listed companies during 2020–2024 through manual and Artificial Intelligence-assisted stages. Tax risk identification performance was assessed using a standardized score based on effective tax rate, book-tax differences, profitability ratios, and cash flow gaps. The findings indicate that Artificial Intelligence-assisted analysis was associated with more consistent, structured, and efficient interpretation of financial statements compared with manual analysis. This study contributes to tax accounting literature by integrating Artificial Intelligence-assisted financial statement analysis into a structured tax risk assessment framework.

KEYWORDS

financial statement analysis; tax risk; artificial intelligence; tax accounting; tax supervision.

Introduction

Identifying tax risks is a key priority in modern tax administration because it is directly linked to the effectiveness of oversight and the sustainability of government revenue (Drobyshevskaya et al., 2020). Tax authorities in various countries have adopted compliance risk management to enhance the efficiency and accuracy of oversight (Strauss et al., 2020). This approach emphasizes the importance of early detection of tax risk signals before conducting intensive audits (Filosa et al., 2025). In this context, the financial statements of publicly listed companies provide a systematic and accessible source of information to support the tax risk identification process. Financial statements reflect a company's profit structure, tax expenses, and financial performance, which can be analyzed to identify patterns relevant to tax risk (Tanko, 2025).

Tax accounting literature indicates that financial statement indicators are correlated with a company's tax position characteristics (Mgammal, 2020). Previous research has used indicators such as the effective tax rate, book-tax differences, profitability ratios, and earnings quality to evaluate discrepancies between accounting profits and tax liabilities (Nissim, 2021). These indicators do not directly indicate non-compliance but provide early signals requiring further evaluation within a risk analysis framework (Benedek & Bognár, 2024). Therefore, financial statement analysis serves as a preventive analytical tool in supporting data-driven tax oversight.

However, the practical application of financial statement indicators still faces various challenges. Analysts often struggle to interpret indicators that are complex and interrelated (Spyromitros & Panagiotidis, 2022). Manual analysis tends to yield varying interpretations due to reliance on individual subjectivity (Dengel et al., 2023). Additionally, the increasing complexity of corporate financial statements and the growing volume of data complicate the process of comprehensive evaluation (Theodorakopoulos et al., 2024). This situation highlights a gap between the availability of financial data and analytical capabilities in consistently identifying tax risks.

Advances in Artificial Intelligence (AI) technology offer opportunities to address these limitations. Digital transformation in the field of accounting is driving the use of analytical technologies to enhance the quality of decision-making (Al-Okaily, 2025). Research in the field of auditing indicates that AI-based technologies can improve the ability to detect patterns and anomalies in financial statements (P. T. Nguyen, 2025)(H. N. Nguyen, 2019). Furthermore, computational approaches enable more systematic information extraction from complex financial statements (Ashtiani & Raahemi, 2021). Thus, Artificial Intelligence has the potential to improve consistency and efficiency in financial statement analysis for the purpose of tax risk identification (Scientific, 2024).

Nevertheless, research integrating financial statement analysis and Artificial Intelligence within the framework of tax risk identification remains limited. Most prior studies have focused on testing the relationship between accounting indicators and tax behaviour using cross-firm quantitative approaches (Cerciello et al., 2023). On the other hand, research on Artificial Intelligence has primarily focused on auditing and fraud detection (Qatawneh, 2025). Research that empirically tests the role of Artificial Intelligence in enhancing the quality of financial statement analysis for tax risk identification remains scarce, particularly in the context of developing countries like Indonesia (Saragih et al., 2023). This gap highlights the need for research that combines both approaches within an integrated analytical framework (Solano & Cruz, 2024).

This study offers several distinct contributions that differentiate it from prior literature. First, this study employs a quasi-experimental design using a difference-in-differences approach to evaluate the causal impact of Artificial Intelligence on analytical performance. This reseach approach is relatively rare in tax accounting research, which typically relies on cross sectional data at the firm level. Second, the study develops a structured Tax Risk Identification Score, which operationalizes financial statement indicators into a consistent analytical framework for assessing the quality of tax risk identification, rather than measuring firm-level tax outcomes. Third, this study positions Artificial Intelligence as a decision-support tool within a tax risk identification framework, distinguishing it from its traditional applications in auditing and fraud detection that focus on anomaly detection. Finally, the study provides empirical evidence in the context of Indonesia as a developing country, where risk-based tax administration and digital transformation initiatives such as Coretax are increasingly relevant, thereby extending the applicability of prior research to emerging economies.

The complexity of tax risk is also increasing due to differences between accounting standards and tax regulations (Neuman et al., 2020). Unlike financial distress risk or fraud risk, tax risk is closely associated with the interpretation of accounting–tax differences, regulatory compliance, and the potential misalignment between financial reporting and taxable income. Tax risk has become increasingly important within the framework of risk-based tax administration because tax authorities rely heavily on financial statement indicators to identify potential compliance risks and prioritize supervision activities. Consequently, tax risk analysis requires not only financial interpretation but also consideration of regulatory and economic contexts.

These differences result in temporary and permanent discrepancies between accounting profit and taxable profit (Görlitz & Dobler, 2023). These differences may reflect legitimate business activities or valid tax planning strategies (Mgammal, 2020). Therefore, identifying tax risk requires a multidimensional approach that considers a combination of various financial statement indicators. Artificial Intelligence can serve as an analytical enhancer that helps identify patterns of relationships among indicators in a systematic and objective manner (How et al., 2020).

Based on these issues, this study formulates two main research questions. First, can financial statement indicators—namely the effective tax rate, book-tax differences, profitability ratios, and cash flow gaps—be used as signals to identify tax risk in publicly listed companies? Second, can the use of Artificial Intelligence improve the accuracy and consistency of tax risk identification compared to manual analysis? These two questions reflect the need to test the effectiveness of the analytical approach.

Methods

This study employs a quasi-experimental research design using a simplified difference-in-differences (DiD) approach to examine changes in tax risk identification performance before and after the implementation of Artificial Intelligence-assisted financial statement analysis (Miller et al., 2020). The DiD framework was applied in a limited and exploratory manner to compare analytical performance between the pre-treatment stage, in which participants conducted manual analysis, and the post-treatment stage, in which participants used Artificial Intelligence-assisted analysis. Since this study did not employ full randomization or a separate control group, the DiD approach is interpreted as a structured comparative framework for evaluating analytical changes rather than as a fully causal econometric estimation.

The study involved 130 Diploma III Tax students consisting of regular and blended learning students who acted as analytical agents in a controlled quasi-experimental setting. Financial statement data were obtained from 25 Indonesian publicly listed companies during the 2020–2024 period. In the pre-treatment stage, participants conducted manual financial statement analysis across multiple sectors, including pharmaceutical, mining, automotive, banking, telecommunications, tobacco, footwear, and food-and-beverage industries, to establish baseline analytical competence under heterogeneous reporting conditions. In the post-treatment stage, participants performed Artificial Intelligence-assisted analysis within a more controlled food-and-beverage sector environment to reduce industry-specific variation and improve comparability.

The experimental intervention consisted of two stages. First, participants received instruction regarding tax-risk identification through financial statement analysis, including the interpretation of effective tax rate, book-tax differences, profitability ratios, and cash flow gaps. Second, participants underwent training in the use of generative Artificial Intelligence for structured financial analysis. The generative Artificial Intelligence system used in this study was ChatGPT based on GPT-4o, accessed through the same web-based interface by all participant groups. To improve procedural consistency, all participants received identical analytical instructions, standardized financial statement inputs, and a uniform prompt structure. The prompts instructed participants to: (1) identify and compute tax-related financial indicators, (2) interpret relationships among indicators and potential tax-risk signals, and (3) classify tax risk into low, medium, or high categories based on the scoring rubric.

To maintain analytical integrity, a Human-in-the-Loop verification procedure was implemented. Participants were required to verify all Artificial Intelligence-generated numerical outputs against the original audited financial statements before submitting the final analysis. The analytical comparison focused on changes in the standardized Tax Risk Identification Score between the pre-treatment and post-treatment stages. The score measured analytical quality based on: (1) accuracy of financial indicator computation, (2) consistency of interpretation, (3) reasoning structure, and (4) appropriateness of tax risk classification. The scoring process employed a standardized analytical rubric applied consistently across both stages of analysis.

The transition from multi-sector analysis in the pre-treatment stage to a controlled food-and-beverage sector analysis in the post-treatment stage was implemented as a methodological control strategy rather than as a change in the primary object of inference. The post-treatment companies were selected using comparability criteria, including availability of complete audited financial statements, similar operating characteristics, comparable cost structures, and sufficient disclosure of tax-related accounts. To minimize task-difficulty bias, both stages applied identical analytical indicators, equivalent reporting periods, standardized scoring rubrics, and uniform analytical instructions. Accordingly, the findings are interpreted as improvements in analytical performance associated with Artificial Intelligence-assisted analysis under the controlled conditions of this study rather than as direct evidence of actual firm-level tax compliance or tax-risk outcomes.

Result and Discussion

The results of the study indicate that financial statement analysis can identify tax risk through key financial indicators, namely the effective tax rate (ETR), book-tax differences (BTD), profitability ratios, and cash flow gaps. During the pre-treatment stage, students analyzed companies across various sectors, resulting in relatively high variability in interpretations. Differences in industry characteristics led to complexity in linking financial indicators to potential tax risks. This situation indicates that manual analysis has limitations in producing consistent interpretations.

The findings indicate that financial statement indicators—namely effective tax rate, book-tax differences, profitability ratios, and cash flow gaps—can function as meaningful signals for tax risk identification when interpreted in combination (J. H. Nguyen, 2021). This supports prior literature that positions financial statement analysis as a tool for detecting tax-related anomalies and risk signals (Velte, 2023).

The statistically significant improvement in post-treatment scores suggests that Artificial Intelligence enhances the consistency and structure of financial analysis. Unlike prior AI applications in auditing and fraud detection, which focus on anomaly detection, this study demonstrates that AI functions as a decision-support tool in interpreting interrelated financial indicators within a tax risk framework.

Furthermore, the absence of significant differences between groups reinforces the internal validity of the quasi-experimental design, indicating that improvements are attributable to the analytical intervention rather than pre-existing differences in participant ability. This aligns with the concept of AI as an analytical enhancer rather than a substitute for professional judgment.

In the context of tax administration, these findings are particularly relevant. Risk-based tax systems require consistent interpretation of financial signals, especially in cases involving related-party transactions, transfer pricing, and profit shifting. The integration of Artificial Intelligence into financial statement analysis may therefore support more systematic and data-driven tax risk identification processes (Artene et al., 2024).

To ensure the reproducibility and objectivity of the Tax Risk Identification Score, this study employed a Standardized Analytical Rubric consisting of four dimensions: (1) accuracy of financial indicator computation, (2) consistency of interpretation across indicators, (3) strength of analytical reasoning, and (4) appropriateness of risk classification. Each dimension was scored on a scale of 0 to 25, resulting in a composite score ranging from 0 to 100. All four dimensions

Table 1. Tax Risks in Public Companies Organized as Business Group

No	Type of Tax Risk	Description
1	Transactions with Related Parties/ Arm's Length Principle (ALP)	Transactions between entities within a business group that do not adhere to the principles of arm's length and business norms
2	Profit Shifting	Shifting profits to entities with lower tax rates or entities incurring losses through unreasonable transactions
3	Thin Capitalization	A financing structure that relies predominantly on debt rather than equity
4	Corporate Actions	Share transfers, mergers, or acquisitions that may affect tax liabilities
5	Double Costing	Charging the same expenses to more than one entity within a business group

Source: Reseacher (processed data)

Table 2. Comparison of Pre- and Post-Treatment Analysis Designs

Aspect	Pre-Treatment	AI-assisted
Method	Manual	AI-assisted
Sector	Multi sector	Single sector (F&B)
Complexity	High	More manageable
Analysis Consistency	Variable	More consistent

Source: Reseacher (processed data)

Table 3. Descriptive Statistics of Tax Risk Identification Scores

Student Group	N	Minimum	Maximum	Mean
Regular	109	76,10	90,00	82,48
Blended Learning	21	70,00	89,00	81,81

Source: Reseacher (processed data)

Table 4. Test of Mean Differences (Independent Samples t-test)

variable	Regular Mean	Blended Mean	T-statistic	p-value
Tax Risk Score	82,48	81,81	0,96	0.34

Source: Reseacher (processed data)

were equally weighted to maintain balance across computational accuracy and analytical judgment.

The "strength of reasoning" dimension was operationalized by assessing the number and coherence of logical relationships established between financial indicators. For example, students were required to demonstrate how variations in book-tax differences (BTD) correspond with changes in effective tax rate (ETR) or profitability measures. Each identified tax risk had to be supported by at least two distinct financial indicators to receive a full score in this dimension. Responses that relied on isolated indicators or lacked logical linkage were assigned lower scores.

The scoring procedure was conducted by two independent evaluators with expertise in tax accounting and financial analysis. To ensure inter-rater reliability, both evaluators applied the rubric independently to a subset of samples, and the level of agreement was assessed using Cohen's kappa coefficient. The resulting kappa value of 0.82 indicates a high level of agreement. Any discrepancies in scoring were subsequently resolved through a third-party adjudication process to ensure consistency and objectivity in the final scores.

Tax risks within the corporate group demonstrate a high level of analytical complexity and serve as the foundation for identifying tax risks through financial statement analysis.

In the post-treatment phase, the analysis focuses on the food and beverage sector, incorporating interventions such as public lectures and the use of Artificial Intelligence. This design change functions as a control mechanism to reduce external variations.

The results indicate that a sector-specific design enhances the stability of the analysis and enables a more focused evaluation (see Table 2).

The relatively balanced mean scores indicate that both groups have comparable baseline abilities, thereby supporting the assumption of baseline comparability (see Table 3).

The test results show no significant difference between the two groups, indicating that improvements in the quality of analysis in the next stage are not influenced by differences in initial ability (see Table 4).

The findings suggest that Artificial Intelligence-assisted analysis was associated with improvements in the consistency, structure, and efficiency of tax risk identification compared with manual analysis (see Table 5). Participants demonstrated stronger ability to integrate multiple financial indicators, particularly in relating effective tax rate, book-tax differences, profitability ratios, and cash flow patterns within a more coherent analytical framework. Compared with the pre-treatment stage, the post-treatment analysis also showed lower interpretive variation across participant groups.

However, these findings should be interpreted within the scope of the controlled quasi-experimental setting. This study evaluates analytical performance in financial statement interpretation rather than actual firm-level tax compliance or professional tax audit outcomes. Accordingly, the results indicate the potential role of Artificial Intelligence as a decision-support mechanism that may support structured tax risk assessment while still requiring professional judgment and human verification. The findings also suggest that Artificial Intelligence-assisted analysis was more effective in identifying quantitative and pattern-based financial relationships, whereas tax issues requiring deeper legal interpretation continued to depend substantially on human analytical evaluation.

The results indicate that the use of Artificial Intelligence improves the quality of financial statement analysis, particularly in integrating multiple financial indicators into a coherent analytical framework. Descriptive and statistical evidence shows that the Tax Risk Identification Score increased from 80.10 in the pre-treatment stage to 82.48 in the post-treatment stage, with a positive post-treatment coefficient ($\beta = 2.32$, $p < 0.01$).

Whereas manual analysis exhibited limitations in systematically correlating indicators such as high Book-Tax Differences (BTD) and anomalous cash flow gaps, the AI-assisted group demonstrated an enhanced capacity to synthesize these signals into structured tax risk assessments. In this context, Artificial Intelligence functions as a form of augmented intelligence, enhancing human analytical capability rather than replacing professional judgment (Jarrahi et al., 2022). This aligns with decision-support theory, which posits that analytical systems improve decision quality by structuring complex information and reducing cognitive bias (Fasolo et al., 2025).

Furthermore, AI-assisted analysis enabled students to process complex disclosures in the notes to financial statements, which are often underutilized in manual analysis. This suggests that AI contributes to expanding the analytical scope beyond primary financial ratios toward more comprehensive financial interpretation. However, the improvement was more pronounced in identifying pattern-based risks, such as transfer pricing signals, which rely on structured relationships among financial indicators. In contrast, risks requiring deeper legal interpretation and regulatory judgment still depended on human validation through a Human-in-the-Loop mechanism. This finding reinforces the role of Artificial Intelligence as a decision-support tool rather than a substitute for professional expertise in tax risk assessment.

Furthermore, the findings of this study become increasingly relevant when linked to tax risks in publicly listed companies organized as business groups, as presented in Table 1. Risks such as related-party transactions that do not meet the principles of arm's length and business norms, profit shifting practices, thin capitalization, corporate actions, and double costing are critical areas in modern tax oversight. In this context, Artificial Intelligence-based financial statement analysis can help identify irregular patterns in a more systematic and data-driven manner.

The findings of this study are relevant to the broader framework of risk-based tax administration, particularly in relation to transfer pricing and related-party transaction analysis. The financial indicators examined in this study,

Table 5. Observed Changes in Tax Risk Identification Performance After Artificial Intelligence-Assisted Analysis

Analytical Aspect	Pre-Treatment (Manual Analysis)	Post-Treatment (AI-Assisted Analysis)	Obeserved Improvement
Consistency of interpretation	Interpretations varied substantially across groups	Interpretations became more standardized across groups	Higher interpretive consistency
Analytical Structure	Financial indicators were often interpreted separately and unsystematically	Financial indicators were interpreted within a more integrated analytical framework	More Systematic analysis
Identification of Indicator Patterns	Limited ability to connect ETR, BTD, profitability and cash flow indicators	Stronger ability to identify relationship among multiple financial indicators	Improved pattern recognition
Analysis Efficiency	Longer time required to review and interpret financial statements	Faster synthesis and interpretation of financial information	Improved analytical efficiency

Source: Reseacher (processed data)

including effective tax rate, book-tax differences, and cash flow patterns, conceptually align with the comparability and economic-substance principles emphasized in Minister of Finance Regulation No. 172 of 2023 concerning the Application of the Arm's Length Principle (ALP) as well as the OECD Transfer Pricing Guidelines (Oguttu, 2020).

However, the present study does not directly evaluate actual transfer-pricing compliance or firm-level tax-risk outcomes. Instead, the study examines analytical performance within a controlled quasi-experimental educational setting. Accordingly, the findings should be interpreted as preliminary evidence regarding the potential usefulness of Artificial Intelligence-assisted financial statement analysis in supporting more structured and data-driven tax risk assessment processes.

Rather than solely confirming compliance expectations, the findings demonstrate how key elements of ALP—such as comparability and economic substance—can be systematically translated into measurable analytical signals, including effective tax rate differentials, book-tax differences, and cash flow inconsistencies. This provides a practical mechanism for identifying transfer pricing risks at an early stage using publicly available financial data.

Furthermore, the results are consistent with the OECD Transfer Pricing Guidelines, which highlight the importance of economic analysis and functional profiling in detecting transfer pricing risks. Importantly, this study contributes to the literature by showing how Artificial Intelligence can enhance the implementation of these principles through augmented analytical processing, enabling more consistent and scalable identification of tax risk. Therefore, the study not only supports existing regulatory frameworks but also advances their application within a data-driven and technology-enabled tax risk identification system.

Thus, the integration of financial statement analysis, Artificial Intelligence, and the tax regulatory framework not only enhances the quality of tax risk identification but also strengthens the practical relevance of this research in supporting risk-based tax oversight, particularly for corporate groups with complex cross-entity transactions.

Overall, the research results indicate that financial statement indicators can serve as tools for identifying tax risks, and the use of Artificial Intelligence enhances the accuracy, consistency, and efficiency of the analysis. These findings address the research questions and reinforce the role of technology in supporting data-driven tax analysis.

The results of this study indicate that financial statement analysis can serve as an initial tool for identifying tax risk through key financial indicators, including the effective tax rate (ETR), book-tax differences (BTD), profitability ratios, and cash flow gaps. During the pre-treatment phase, the analysis was conducted across companies in various sectors, resulting in a relatively high degree of interpretive variation. This variation reflects the complexity of industry characteristics that influence financial statement structure and potential tax risks (Neuman et al., 2020). These findings are consistent with prior literature indicating that financial indicators capture underlying tax-related characteristics of firms but require careful interpretation due to the influence of economic conditions and accounting policies (Allen et al., 2021)

From a theoretical perspective, these findings suggest that tax risk identification is not merely a function of isolated financial indicators, but rather depends on the analytical integration of multiple interrelated signals. This aligns with the broader view in tax accounting that discrepancies between accounting income and taxable income, as reflected in BTD and ETR variations, represent potential indicators of tax planning behaviour rather than direct evidence of non-compliance. The observed variability in manual interpretation highlights the role of cognitive limitations and interpretive subjectivity in financial analysis, reinforcing the need for structured analytical frameworks in tax risk assessment.

Importantly, these results matter specifically for tax risk analysis because tax authorities do not rely on single indicators but on patterns of financial signals to identify potential compliance risks. The findings of this study provide experimental evidence that the identification of such patterns is inherently complex and prone to inconsistency under manual analysis. Therefore, the contribution of this study lies in demonstrating that financial statement analysis can function as an early-stage risk screening mechanism, rather than a definitive measure of tax risk.

It is important to distinguish between the empirical findings and broader policy implications. The experimental results provide evidence that analytical performance improves under structured conditions; however, the study does not directly measure firm-level tax risk or validate actual tax compliance outcomes. The implications for tax administration—such as the potential integration of data-driven analysis into risk-based supervision systems—should therefore be interpreted as contextual extensions of the findings rather than direct empirical conclusions.

In the post-treatment phase, the analysis focused on the food and beverage sector, accompanied by interventions in the form of a lecture and the use of Artificial Intelligence. The results showed that students were able to produce more consistent and systematic analyses. The average Tax Risk Identification Score for regular students was 82.48 and for blended learning students was 81.81, indicating that both groups had relatively comparable initial abilities. The non-significant results of the mean difference test support the assumption of baseline comparability, so that the improvement in the quality of analysis can be attributed to the interventions provided.

The use of Artificial Intelligence was associated with improvement the quality of tax risk identification. Students were able to integrate various financial indicators into a more comprehensive analytical framework. Artificial Intelligence assists in identifying patterns that are difficult to detect through manual analysis, particularly in linking ETR, BTD, and cash flow. These findings are consistent with the research by (El-Feel et al., 2025), which shows that analytical technology enhances the ability to detect patterns in financial data, as well as Moll and (Hung et al., 2023), who affirm that digitalization improves the quality of decision-making in accounting.

In the context of group companies, the results of this study are highly relevant. Tax risks such as related-party transactions that do not meet the principles of arm's length and business norms, profit shifting practices, thin capitalization, corporate actions, and double costing are key areas in tax oversight. AI-based analysis enables the systematic identification of these risks through financial patterns reflected in financial statements. This aligns with the provisions in PMK 172/PMK.03/2023, which emphasizes the importance of comparability analysis and related-party transactions in determining the reasonableness of transfer pricing.

Furthermore, the findings of this study also align with the OECD Transfer Pricing Guidelines (OECD, 2022), which state that economic and functional analysis form the basis for identifying transfer pricing risks and potential profit shifting. The OECD emphasizes the importance of a data-driven approach in evaluating the reasonableness of transactions between entities within a business group. Thus, the use of Artificial Intelligence can be viewed as a tool that strengthens the implementation of these principles in practice.

From a methodological perspective, the use of the difference-in-differences approach makes a significant contribution to this study. This approach allows for the isolation of the intervention's effects from general learning effects, thereby providing stronger empirical evidence regarding the causal relationship between the use of Artificial Intelligence and improved analysis quality. These findings expand upon previous literature, which generally employed cross-sectional approaches in tax analysis (Mohammed & Tangl, 2023).

Despite the significant improvements observed, this study identifies a critical boundary in AI performance: the risk of 'Algorithm Over-reliance.' While AI excelled at identifying quantitative patterns such as Thin Capitalization or ETR anomalies, it occasionally struggled with context-heavy 'Corporate Actions' that required deep qualitative legal interpretation. Furthermore, in 15% of the cases, participants noted that the AI attempted to 'hallucinate' tax regulations that were not applicable to the Indonesian context. This highlights that while AI serves as a powerful analytical enhancer, its role in tax oversight must remain a collaborative one, where professional skepticism and human judgment provide the final layer of validation.

From a methodological perspective, the use of the difference-in-differences (DiD) approach represents a key contribution of this study. The DiD framework is particularly suitable in this context because it allows the separation of improvements in analytical performance attributable to the Artificial Intelligence intervention from general learning effects that naturally occur over time. In a setting where the same participants are exposed to sequential analytical tasks, conventional pre–post comparisons risk conflating skill development with treatment effects. By incorporating both temporal variation (pre- and post-treatment) and group comparison, the DiD approach provides a more robust identification strategy for isolating the incremental impact of AI-assisted analysis. This is especially relevant in evaluating human–AI interaction, where improvements may arise from both technological augmentation and experiential learning. Therefore, the application of DiD strengthens the internal validity of the study and extends prior tax research, which has predominantly relied on cross-sectional designs (Mohammed & Tangl, 2023).

Despite the observed improvements in analytical performance, this study also identifies important limitations in the use of Artificial Intelligence. Empirical observations from the scoring process indicate that errors occurred in three primary forms. First, numerical extraction errors, where AI-generated values did not precisely match the reported figures in the financial statements. Second, interpretive errors, where the relationships between financial indicators—such as the linkage between book-tax differences and effective tax rate—were identified but not consistently interpreted within a coherent tax risk framework. Third, contextual hallucinations, where the AI generated references to tax regulations that were not applicable within the Indonesian tax system.

These findings suggest that while Artificial Intelligence functions as a powerful analytical enhancer, its outputs are not fully reliable without human validation. Accordingly, the role of AI in tax risk identification should be understood within a Human-in-the-Loop framework, where professional judgment and skepticism serve as the final layer of verification. This reinforces the position of Artificial Intelligence as a decision-support tool rather than a substitute for expert evaluation in tax analysis.

Overall, the results of this study indicate that financial statement indicators can serve as early signals in identifying tax risk, and the use of Artificial Intelligence enhances the quality, consistency, and efficiency of analysis (Rahman et al., 2024). The integration of financial statement analysis and Artificial Intelligence makes a significant contribution to strengthening data-driven tax analysis.

Conclusion

This study examines whether financial statement indicators can function as signals for identifying tax risk and whether Artificial Intelligence-assisted analysis improves the quality of tax risk identification compared with manual analysis. Based on the findings, financial indicators such as effective tax rate, book-tax differences, profitability ratios, and cash flow gaps can serve as meaningful analytical signals for identifying potential tax risk, particularly when interpreted within an integrated analytical framework rather than in isolation. The findings also indicate that Artificial Intelligence-assisted analysis was associated with more consistent, structured, and efficient interpretation of financial statements compared with manual analysis under the conditions of this study. Accordingly, the study suggests that Artificial Intelligence may support tax risk assessment as a decision-support mechanism while professional judgment and human verification remain essential.

From a practical perspective, these findings provide preliminary implications for the development of risk-based tax supervision within the Directorate General of Taxes (DGT), particularly in supporting more structured financial statement analysis and early-stage tax risk screening. The study also highlights the importance of Artificial Intelligence literacy and analytical capability development in tax education to prepare future tax professionals for increasingly data-driven tax administration environments. These findings are also consistent with recent literature in accounting and taxation, which suggests that Artificial Intelligence may improve analytical consistency and efficiency in financial analysis. However, the findings of this study also indicate that Artificial Intelligence-assisted analysis should remain complemented by professional judgment and human verification, particularly in interpreting complex tax-related issues.

This study contributes to tax accounting literature by integrating financial statement analysis and Artificial Intelligence within a structured quasi-experimental framework for tax risk assessment. However, several limitations should be acknowledged. First, the study uses Diploma III Tax students as proxy analytical agents rather than professional tax auditors or practitioners. Second, the analysis was conducted within a controlled quasi-experimental educational setting that may not fully reflect real-world tax administration practices. Third, the pre-treatment and post-treatment stages did not necessarily use identical company samples, although comparability controls were applied to reduce sectoral bias. Finally, this study evaluates analytical quality in tax risk identification rather than actual firm-level tax compliance or transfer-pricing outcomes. Therefore, future research is recommended to incorporate practitioner-based analysis, actual tax audit data, and broader sectoral coverage to improve the generalizability and practical applicability of Artificial Intelligence-assisted tax risk assessment models.

Author contributions

Supriyadi conceived and designed the research framework, developed the quasi-experimental design, supervised the implementation of the financial statement analysis activities, interpreted the empirical findings, and drafted the manuscript, particularly the introduction, methodology, results, and discussion sections.

Arief Budi Wardana contributed to data collection, rubric development, statistical analysis, and validation of the Tax Risk Identification Score. He also assisted in standardizing the Artificial Intelligence prompting procedures and verifying the analytical outputs generated during the experiment.

IGK Chaya Bayu A provided academic supervision in tax accounting, financial statement analysis, and research methodology. He also contributed to critical review, language refinement, and manuscript revision to ensure academic quality, coherence, and compliance with publication standards.

Nasikhudin contributed from the perspective of the Directorate General of Taxes (DGT), particularly in relation to risk-based taxpayer supervision and the utilization of Artificial Intelligence for taxpayer risk mapping. He also contributed to the refinement and finalization of the manuscript to strengthen the alignment between the theoretical framework, analytical findings, and their practical relevance within tax administration.

All authors are responsible for the content of this manuscript and have reviewed and approved the final version.

Funding

This research was conducted independently using the authors' personal resources without financial support from any governmental, institutional, commercial, or non-profit funding agency. All stages of the study, including research design, data collection, analysis, interpretation, and manuscript preparation, were carried out independently by the authors. The authors confirm that no external party influenced the conduct or outcomes of this research.

Acknowledgements

The authors would like to express sincere appreciation to the Indonesia Stock Exchange (IDX) and the official websites of publicly listed companies for providing access to the annual reports and financial statement data used in this study. The authors are also grateful to guest lecturers, tax practitioners, tax auditors, academic colleagues, and Diploma III Tax students who contributed valuable insights and participated in the quasi-experimental financial statement analysis activities. In addition, the authors appreciate all individuals who provided constructive feedback during the preparation of this manuscript.

Conflict of interest

The authors declare that there are no professional, financial, institutional, or personal conflicts of interest that could influence the objectivity, independence, or interpretation of this study. The study was conducted independently and without undue external influence. This statement reflects the authors' commitment to academic integrity, transparency, and research credibility.

References

Allen, E., O'Leary, D. E., Qu, H., & Swenson, C. W. (2021). Tax specific versus generic accounting-based textual analysis and the relationship with effective tax rates: Building context. *Journal of Information Systems*, *35*(2), 115–147.

Al-Okaily, M. (2025). Attitudes toward the adoption of accounting analytics technology in the digital transformation landscape. *Journal of Accounting & Organizational Change*, *21*(3), 593–613. https://doi.org/https://doi.org/10.1108/JAOC-04-2024-0127

Artene, A. E., Domil, A. E., & Ivascu, L. (2024). Unlocking business value: Integrating AI-driven decision-making in financial reporting systems. *Electronics*, *13*(15), 3069.

Ashtiani, M. N., & Raahemi, B. (2021). Intelligent fraud detection in financial statements using machine learning and data mining: a systematic literature review. *Ieee Access*, *10*, 72504–72525.

Benedek, P., & Bognár, F. (2024). *Compliance Risk Assessment–Results of a Comprehensive Literature Revie*.

Cerciello, M., Busato, F., & Taddeo, S. (2023). The effect of sustainable business practices on profitability. Accounting for strategic disclosure. *Corporate Social Responsibility and Environmental Management*, *30*(2), 802–819.

Dengel, A., Gehrlein, R., Fernes, D., Görlich, S., Maurer, J., Pham, H. H., Großmann, G., & Eisermann, N. D. genannt. (2023). Qualitative research methods for large language models: Conducting semi-structured interviews with ChatGPT and BARD on computer science education. *Informatics*, *10*(4), 78.

Drobyshevskaya, L., Vylegzhanina, E., Grebennikova, V., & Mamiy, E. (2020). The main approaches to assessing efficiency of tax administration and control in the context of digitalization. *International Conference on Integrated Science*, 95–111.

El-Feel, H. W. T., Amin, H. M. G., Mohamed, D. M., & Mohamed, E. K. A. (2025). Assessing board attributes in shaping corporate tax behavior: a bibliometric analysis and implications for future research. *Journal of Accounting Literature*, 1–42.

Fasolo, B., Heard, C., & Scopelliti, I. (2025). Mitigating cognitive bias to improve organizational decisions: An integrative review, framework, and research agenda. *Journal of Management*, *51*(6), 2182–2211.

Filosa, J. R., Huang, J., Lei, L., & Stein, S. E. (2025). Does tax enforcement inform auditors' risk assessment? Evidence from key audit matters. *Contemporary Accounting Research*, *42*(2), 1423–1454.

Görlitz, A., & Dobler, M. (2023). Financial accounting for deferred taxes: A systematic review of empirical evidence. *Management Review Quarterly*, *73*(1), 113–165.

How, M.-L., Cheah, S.-M., Chan, Y.-J., Khor, A. C., & Say, E. M. P. (2020). Artificial intelligence-enhanced decision support for informing global sustainable development: A human-centric AI-thinking approach. *Information*, *11*(1), 39.

Hung, B. Q., Hoa, T. A., Hoai, T. T., & Nguyen, N. P. (2023). Advancement of cloud-based accounting effectiveness, decision-making quality, and firm performance through digital transformation and digital leadership: Empirical evidence from Vietnam. *Heliyon*, *9*(6).

Jarrahi, M. H., Lutz, C., & Newlands, G. (2022). Artificial intelligence, human intelligence and hybrid intelligence based on mutual augmentation. *Big Data & Society*, *9*(2), 20539517221142824.

Mgammal, M. H. (2020). Corporate tax planning and corporate tax disclosure. *Meditari Accountancy Research*, *28*(2), 327–364.

Miller, C. J., Smith, S. N., & Pugatch, M. (2020). Experimental and quasi-experimental designs in implementation research. *Psychiatry Research*, *283*, 112452.

Mohammed, H., & Tangl, A. (2023). Taxation perspectives: Analyzing the factors behind viewing taxes as punishment—A comprehensive study of taxes as service or strain. *Journal of Risk and Financial Management*, *17*(1), 5.

Neuman, S. S., Omer, T. C., & Schmidt, A. P. (2020). Assessing tax risk: Practitioner perspectives. *Contemporary Accounting Research*, *37*(3), 1788–1827.

Nguyen, H. N. (2019). Enhancing the capacity of tax authorities and its impact on transfer pricing activities of FDI enterprises in Ha Noi, Ho Chi Minh, Dong Nai, and Binh Duong province of Vietnam. *Management Science Letters*, *9*(8), 1299–1310. https://doi.org/10.5267/j.msl.2019.4.011

Nguyen, J. H. (2021). Tax avoidance and financial statement readability. *European Accounting Review*, *30*(5), 1043–1066.

Nguyen, P. T. (2025). AI Technology in Auditing and Financial Error Detection. *International Congress on Information and Communication Technology*, 89–105.

Nissim, D. (2021). Earnings quality. *Columbia Business School Research Paper Forthcoming. Assessed from Https://Ssrn. Com/Abstract*, *3794378*.

OECD. (2022). OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations 2022. In *OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations*. OECD.

Oguttu, A. W. (2020). Challenges of applying the comparability analysis in curtailing transfer pricing: Evaluating the suitability of some alternative approaches in Africa. *Intertax*, *48*(1).

Qatawneh, A. M. (2025). The role of artificial intelligence in auditing and fraud detection in accounting information systems: moderating role of natural language processing. *International Journal of Organizational Analysis*, *33*(6), 1391–1409.

Rahman, S., Sirazy, M. R. M., Das, R., & Khan, R. S. (2024). An exploration of artificial intelligence techniques for optimizing tax compliance, fraud detection, and revenue collection in modern tax administrations. *International Journal of Business Intelligence and Big Data Analytics*, *7*(3), 56–80.

Saragih, A. H., Reyhani, Q., Setyowati, M. S., & Hendrawan, A. (2023). The potential of an artificial intelligence (AI) application for the tax administration system's modernization: the case of Indonesia. *Artificial Intelligence and Law*, *31*(3), 491–514.

Scientific, L. L. (2024). Optimization of smart taxation using artificial intelligence: risks and opportunities. *Journal of Theoretical and Applied Information Technology*, *102*(5).

Solano, M. C., & Cruz, J. C. (2024). Integrating analytics in enterprise systems: A systematic literature review of impacts and innovations. *Administrative Sciences*, *14*(7), 138.

Spyromitros, E., & Panagiotidis, M. (2022). The impact of corruption on economic growth in developing countries and a comparative analysis of corruption measurement indicators. *Cogent Economics & Finance*, *10*(1), 2129368.

Strauss, H., Fawcett, T., & Schutte, D. (2020). Tax risk assessment and assurance reform in response to the digitalised economy. *Journal of Telecommunications and the Digital Economy*, *8*(4), 96–126.

Tanko, U. M. (2025). Financial attributes and corporate tax planning of listed manufacturing firms in Nigeria: moderating role of real earnings management. *Journal of Financial Reporting and Accounting*, *23*(3), 1024–1056.

Theodorakopoulos, L., Thanasas, G., & Halkiopoulos, C. (2024). Implications of big data in accounting: Challenges and opportunities. *Emerging Science Journal*, *8*(3), 1201–1214.

Velte, P. (2023). The link between corporate governance and corporate financial misconduct. A review of archival studies and implications for future research. *Management Review Quarterly*, *73*(1), 353–411.