

An Ethical Audit Model for the Governance of Indonesia's Digital Tax System: Integrating the Principles of Fairness, Transparency, and Accountability

Rita Nataliawati¹, Qosim², Tri Winarsih³, Sri Yaumi⁴

Faculty of Economics and Business, Institut Teknologi dan Bisnis Ahmad Dahlan Lamongan

Email: rita12natalia@gmail.com, qosim@ahmaddahlan.ac.id, triwinarsih3113@gmail.com, sriyaumiitbadla@gmail.com

Keywords:

Ethics Audit,
Artificial
Intelligence,
Digital Tax
Governance,
Fairness,
Accountability.

Abstract

The transformation of the digital tax system through the implementation of the Core Tax Administration System (CTAS) marks a new era in artificial intelligence (AI) based fiscal governance in Indonesia. While bringing high efficiency, AI implementation also raises ethical challenges such as algorithmic bias, lack of transparency in automated decisions and unclear human responsibility for system outcomes. This study aims to develop an AI Ethics Audit Model that integrates the principles of fairness, transparency and accountability (FTA) to strengthen fair and accountable digital tax governance. The research approach used is a qualitative descriptive approach with a development research model design, which combines a literature review of 45 journals (2018–2025), in-depth interviews with five experts (three DGT officials and two AI ethics academics), and analysis of policy documents such as PMK No. 81 of 2024 and ISO/IEC 42001:2023. The results show that there is still a risk of data bias, limited algorithm documentation and the absence of a formal ethics oversight mechanism at the DGT. Based on these findings, five stages of an AI ethics audit are formulated: audit planning, ethical risk identification, control evaluation, audit reporting and monitoring and follow up. This model provides a theoretical contribution by strengthening the concept of ethical assurance in digital governance, and a practical contribution by providing an ethical oversight framework for the Directorate General of Taxes (DGT) to ensure the transparent, fair and responsible implementation of AI in the national tax system.

1. Introduction

The rapid advancement of artificial intelligence (AI) technologies has profoundly transformed public sector governance, including tax administration systems (Sawitri, 2025). In Indonesia, the digitalization of taxation through the implementation of the Core Tax Administration System (CTAS) by the Directorate General of Taxes (DJP) represents a major structural reform towards data-driven and technology-enabled governance. CTAS enables automation of administrative processes, enhanced risk-based compliance analytics, and integrated cross-platform tax data management, all of which are designed to improve the efficiency and effectiveness of tax collection (Ministry of Finance, 2024). Despite these promising developments, the increasing reliance on AI also introduces complex ethical and systemic challenges. AI-based decision-making may create risks of algorithmic bias, opacity in automated processes, and blurred

lines of accountability between human officers and autonomous systems (Floridi & Cows, 2022). In the tax domain, biased or incomplete data may result in unfair compliance assessments or disproportionate audit prioritization, which can ultimately undermine public trust in tax authorities (Jobin et al., 2019).

A growing body of literature has examined the ethical implications of AI and its governance in the public sector. Floridi and Cows (2022) emphasize fairness, transparency, and accountability (FTA) as the foundational global principles of ethical AI. Rahwan (2017) proposes the society-in-the-loop framework, stressing the need for continuous human oversight in autonomous decision systems. The OECD (2021) further recommends embedding algorithmic accountability within AI-enabled public sector applications, including taxation. However, most studies remain conceptual and do not extend to applied audit mechanisms or ethical assurance frameworks tailored to the

context of digital tax administration. In Indonesia, there is no existing ethical audit model specifically designed to evaluate AI systems within the DJP, especially for the CTAS environment. Yet ethical assurance is crucial to ensuring that AI-enabled government systems uphold fairness, openness, and public responsibility (Sætra, 2023). Studies such as Cath (2018) and Bryson et al. (2020) have introduced ethical approaches for AI governance in the public sector; however, they primarily focus on regulatory perspectives rather than the development of comprehensive audit models.

This gap demonstrates a significant disconnection between normative ethical AI principles and their practical application in technical oversight mechanisms, particularly in the taxation sector. Existing research predominantly addresses governance structures and broad ethical guidelines for public-sector AI. Nevertheless, studies offering practical and implementable ethical audit models for digital taxation systems remain limited, especially in the Indonesian context. Therefore, this study seeks to develop an AI Ethical Audit Model that systematically integrates fairness, transparency, and accountability principles to strengthen the governance of digital tax systems.

Fairness emphasizes the need to mitigate algorithmic bias and ensure equitable treatment of taxpayers. Transparency highlights the importance of explainability and auditability of AI-based decisions to enhance public trust and enable responsible oversight. Accountability ensures that human officers retain ultimate responsibility through ethical documentation, monitoring, and decision-review mechanisms. Together, these principles form the ethical foundation required to ensure that CTAS not only delivers efficiency but also upholds social justice and taxpayer trust.

Accordingly, the objective of this study is to formulate an integrated AI Ethical Audit Model grounded in the principles of fairness, transparency, and accountability to support the development of a just, transparent, and

accountable digital tax governance system in Indonesia. The proposed model is expected to contribute to strengthening ethical compliance, improving system integrity, and enhancing public confidence in the national taxation framework.

2. Literature Review

2.1 Agency Theory and Stakeholder Theory

Agency Theory, as introduced by Jensen and Meckling (1976), describes the contractual relationship between principals and agents. In the context of taxation, citizens function as principals while the government acts as the agent responsible for managing public revenue. The digitalization of tax administration introduces AI as a new intermediary within this agency relationship, where certain decision-making processes traditionally executed by human officers are now automated by intelligent systems. This shift also transforms the role of auditors from assessing transactional compliance to supervising the integrity and ethical reliability of digital tax systems (Arens et al., 2021).

Conversely, Stakeholder Theory (Freeman, 1984) emphasizes that public policy, including taxation, must consider the interests of all stakeholders—citizens, government institutions, and the business sector. The development of an AI Ethical Audit Model therefore becomes an important mechanism for balancing technological efficiency with social justice. Together, these theories highlight the need for new accountability structures to ensure that the integration of AI in tax administration maintains transparency, fairness, and public trust.

2.2 Artificial Intelligence Ethics

Ethical considerations in AI have become increasingly prominent in the past decade as algorithmic systems play a larger role in public-sector decision-making. Floridi and Cowls (2022) propose fairness, transparency, and accountability (FTA) as foundational principles for ethical AI governance. These principles serve not only as moral guidelines

but also as operational instruments for evaluating, auditing, and regulating intelligent systems. Jobin et al. (2019), in their review of more than 80 global AI ethics guidelines, identify a substantial gap between high-level ethical principles and real-world implementation. Many institutions articulate ethical standards yet lack formal mechanisms for monitoring and enforcing algorithmic compliance. Consequently, AI ethical auditing has emerged as an essential component of responsible public-sector digital governance (Cath, 2018).

Bryson et al. (2020) argue that public-sector AI requires structured ethical audit mechanisms to ensure that automated decisions do not violate principles of justice or human rights. Such mechanisms may include assessments of algorithmic fairness, transparency testing, and the establishment of AI ethics oversight bodies. Despite growing interest internationally, research in Indonesia has not fully explored the linkage between AI ethics and oversight of digital tax administration systems. This gap underscores the need for a comprehensive audit framework tailored to the Indonesian taxation environment.

2.3 Digital Tax Governance

Indonesia's transition toward digital tax governance is marked by the introduction of the Core Tax Administration System (CTAS), designed to create an integrated, efficient, and data-driven tax administration environment. This reform aligns with OECD (2023a) recommendations urging member states to adopt data-driven compliance models and automated tax processes. However, the OECD (2021) also stresses that AI accountability and algorithmic transparency must accompany such modernization efforts to avoid reinforcing existing inequities or generating new forms of digital discrimination.

AI ethical auditing thus becomes a critical oversight layer beyond traditional financial or compliance audits. Rahwan (2017) introduces the society-in-the-loop concept,

which calls for the involvement of the public and independent auditors in overseeing algorithmic decision-making in the public sector. This approach ensures that AI-driven decisions remain accountable and socially grounded. In Indonesia, however, most studies still focus on the technical implementation of CTAS rather than its ethical or algorithmic oversight dimensions. Therefore, developing an AI Ethical Audit Model grounded in fairness, transparency, and accountability is essential for closing this gap and ensuring equitable digital tax governance.

2.4 Fairness in Algorithmic Decision-Making

Fairness is a central ethical principle requiring that AI-driven decision processes be free from discriminatory bias. Within the taxation domain, algorithmic fairness implies that AI systems used by the Directorate General of Taxes (DJP) must not generate outcomes that disproportionately disadvantage certain groups of taxpayers. Mehrabi et al. (2021) identify sources of algorithmic bias including imbalanced training datasets, misclassification errors, and data representations that do not adequately capture the characteristics of real-world populations.

To operationalize fairness within an AI ethical audit, the following elements are essential:

1. **Dataset and labeling evaluation** to ensure the absence of discrimination based on geography, business type, or economic capacity.
2. **Algorithmic fairness testing**, particularly through indicators such as demographic parity and equal opportunity, to detect potential bias.
3. **Corrective policy recommendations** when disparities arise between taxpayer groups with similar profiles but differing algorithmic outcomes.

Fairness-driven audits play a crucial role in strengthening taxpayer trust, reinforcing the social justice principles embedded within Indonesia's tax system, and ensuring that AI-driven tax assessments remain equitable.

2.5 Transparency in AI Systems

Transparency refers to the ability of an AI system to explain the rationale behind its decisions. Within CTAS, advanced analytics and AI-driven models are utilized for tax risk profiling, compliance detection, and behavior prediction. Without adequate transparency, these processes risk generating public perceptions of opacity or injustice, which may undermine institutional legitimacy. Doshi-Velez and Kim (2018) conceptualize transparency in terms of two components: **explainability**, which pertains to human-understandable explanations of system operations; and **auditability**, which enables external verification of system outputs.

In the context of AI ethical auditing, transparency can be operationalized through:

1. **Comprehensive documentation** of algorithmic processes including model parameters, training methods, and data sources.
2. **Digital audit trails** capturing every automated decision for use in internal or external audits.
3. **Integration of Explainable AI (XAI) tools** within CTAS to allow auditors and policymakers to interpret system logic and assess decision validity.

Enhancing transparency ensures not only internal clarity within DJP but also bolsters public confidence that AI-enabled tax decisions are objective and verifiable.

2.6 Accountability in Digital Tax Administration

Accountability emphasizes the necessity of clearly defining human responsibility throughout the lifecycle of AI-driven decision-making. The OECD (2021) asserts that accountability mechanisms are essential to ensure that AI remains under human oversight and complies with legal and ethical standards. In Indonesia's tax context, accountability extends beyond system developers to include DJP officials who serve as decision owners for AI-generated outcomes. AI

ethical auditing assesses the robustness of the organizational structure, internal controls, and reporting systems that enforce accountability. Key components of accountability implementation within an ethical audit model include:

1. **Responsibility mapping** to delineate roles among system developers, system operators, and decision-makers.
2. **Establishment of an AI Ethics Oversight Board** within DJP to supervise algorithmic governance.
3. **Periodic ethical audit reporting** as part of digital governance performance assessment.

Accountability thus functions as the connective principle ensuring that AI-driven decisions remain lawful, ethical, and attributable to responsible human authorities.

3. Research Methods

3.1 Research Design

This study adopts a descriptive qualitative approach with a model development research design aimed at constructing and validating an AI Ethical Audit Model for digital tax governance in Indonesia. The qualitative approach is appropriate because it enables an in-depth exploration of ethical principles, audit mechanisms, and the operational dynamics of AI in tax administration, which cannot be quantified or captured through statistical measurement (Creswell & Poth, 2018). The model development orientation supports the iterative formulation, refinement, and validation of a conceptual audit framework aligned with international standards for AI governance.

3.2 Data Collection Techniques

Data were collected using three complementary methods to ensure methodological triangulation and enhance the robustness of findings.

3.2.1 Literature Review

A comprehensive review of 45 national and international journal articles published between 2018 and 2025 was conducted. The

literature review focused on key themes including AI ethics, algorithmic governance, digital tax systems, accountability mechanisms, and audit frameworks. This step provided a theoretical foundation and enabled the identification of global best practices relevant to AI ethical assurance.

3.2.2 Expert Interviews

Semi-structured interviews were conducted following the qualitative inquiry approach suggested by Yin (2018). Participants included academics, AI practitioners, tax officials, and digital governance experts. The interviews aimed to capture professional insights regarding the ethical challenges, risks, and auditing needs associated with AI-enabled tax administration. Interview guides were developed to ensure consistency while allowing flexibility for probing emergent themes.

3.2.3 Document Analysis

Document analysis was conducted to examine regulatory, institutional, and technical frameworks influencing AI implementation in tax administration. Key documents analyzed include:

1. Ministry of Finance Regulation PMK No. 81/2024 on CTAS,
2. Law No. 7/2021 on the Harmonization of Tax Regulations (HPP),
3. ISO/IEC 42001:2023 on Artificial Intelligence Management Systems (ISO, 2023).

These documents provided essential context regarding digital tax governance, ethical requirements, and compliance expectations.

3.3 Data Analysis

The collected data were analyzed using the interactive model proposed by Nataliawati et al. (2024), consisting of three iterative stages:

3.3.1 Data Reduction

This stage involved categorizing and simplifying raw data by identifying key themes associated with fairness, transparency, and accountability. Data were coded and grouped to

highlight recurring patterns relevant to the development of the AI Ethical Audit Model.

3.3.2 Data Display

Organized data were presented in matrices, narrative descriptions, and conceptual diagrams to facilitate interpretation. This step enabled thematic comparison and the identification of relationships between categories.

3.3.3 Conclusion Drawing and Verification

In this phase, analytical conclusions were formulated and cross-checked through triangulation between literature sources, interview findings, and regulatory documents. Verification ensured that the emerging audit model was theoretically grounded, contextually relevant, and logically coherent.

3.4 Ethical Considerations

Ethical principles were adhered to throughout the research process. Confidentiality of informants was strictly maintained, and all collected data were used exclusively for academic purposes. Participants were provided with clear information regarding the study objectives, and informed consent was obtained prior to conducting interviews. Additionally, participants were granted the right to review and validate the interpretation of their statements before inclusion in the final analysis (Neuman, 2014). These ethical safeguards enhanced the credibility, transparency, and trustworthiness of the research findings.

4. Results and Discussion

4.1 Research Results

The empirical findings of this study were generated through an integrated analysis of in-depth interviews and systematic literature review. The informants consisted of five key experts: three officials from the Directorate General of Taxes (DJP) involved in digital transformation and two academic specialists in AI ethics. All participants possessed more than ten years of professional experience and had

direct involvement in the development, implementation, or evaluation of the Core Tax Administration System (CTAS).

Overall, the results indicate that AI deployment within Indonesia's digital tax administration has provided substantial benefits, particularly in improving administrative efficiency, strengthening compliance analytics, and enhancing data quality. However, several ethical concerns emerged consistently across interviews. These issues include:

1. **algorithmic bias** attributable to imbalanced training datasets,
2. **insufficient transparency** due to limited documentation of AI decision-making processes, and

3. **unclear accountability mechanisms** for decisions produced by automated systems.

Based on thematic analysis, this study developed the **AI Ethical Audit Model**, a structured framework consisting of five core stages designed to oversee ethical compliance in tax-based AI systems:

- (1) Ethical audit planning, identification of ethical risks,
- (2) Evaluation of ethical control mechanisms,
- (3) Ethical audit reporting, and
- (4) Follow-up monitoring and continuous improvement.

Table 1. Summary of Key Findings on the Integration of FTA Principles in the Digital Tax System

No	Ethical Principle	Field Findings	Key Recommendations
1	Fairness	Presence of bias in datasets for small and medium taxpayers due to domination of large corporate data.	Diversify datasets and conduct periodic bias testing.
2	Transparency	Limited documentation of algorithms and audit trails in CTAS.	Implement Explainable AI (XAI) and automated audit logs.
3	Accountability	Absence of a dedicated AI ethics oversight unit in DJP.	Establish an AI Ethics Oversight Board and ethical responsibility policy.

Source: In-depth Interviews (2025)

The findings further suggest strong institutional support for establishing a formal ethical audit mechanism within DJP. The average agreement score (1–5 scale) on the importance of implementing an AI ethics audit framework reached **4.6**, reflecting a high level of perceived urgency and relevance of the proposed model to future digital governance.

4.2 Research Discussion

The findings of this study confirm and extend existing evidence that the integration of artificial intelligence into public tax administration must be accompanied by a robust ethical oversight system. The proposed **AI Ethical Audit Model** contributes a structured governance mechanism grounded in

the principles of fairness, transparency, and accountability (FTA), which are widely recognized as the cornerstone of trustworthy AI in the public sector.

Fairness Dimension

The fairness component of the model ensures that algorithmic-based tax decisions do not disproportionately disadvantage specific groups of taxpayers. This aligns with the argument of **Mehrabi et al. (2021)**, who emphasize that algorithmic fairness requires continuous monitoring of data representativeness and the reduction of bias arising from skewed datasets. In the context of CTAS, fairness auditing involves evaluating training dataset composition and integrating

automated bias detection tools to mitigate discriminatory outcomes. These mechanisms are essential in preventing systemic inequities in tax compliance analysis.

Transparency Dimension

Transparency emerged as a critical concern in the field findings due to the limited visibility of AI decision-making processes. The recommendation to integrate Explainable Artificial Intelligence (XAI) is consistent with the perspective of Doshi-Velez and Kim (2018), who argue that interpretability is crucial for both auditability and public accountability of AI systems. Increasing algorithmic transparency enables tax auditors, policymakers, and oversight bodies to understand and validate AI-generated decisions. Moreover, enhanced transparency closely aligns with broader public sector good governance principles as highlighted by the OECD (2023), particularly regarding openness and traceability of automated decision-making.

Accountability Dimension

The accountability principle within the proposed model clarifies the division of responsibility between automated systems and human decision-makers. As noted by Sætra (2023), accountability cannot be fully delegated to AI systems; instead, institutions must establish formal mechanisms that ensure human oversight, ethical responsibility, and remediation pathways. Accordingly, this study recommends the formation of an AI Ethics Oversight Board within DJP, responsible for supervising AI operations, coordinating ethical audits, and ensuring compliance with national regulations and international standards such as ISO/IEC 42001:2023. This governance structure strengthens institutional readiness to manage ethical risks associated with emerging AI-driven tax processes.

Overall, the integration of the FTA principles within the proposed AI Ethical Audit Model contributes a novel framework for addressing ethical vulnerabilities in digital tax

administration. The findings demonstrate that ethical governance must evolve concurrently with technological adoption to ensure responsible, equitable, and transparent implementation of AI within public institutions.

5. Closing

5.1 Conclusion

This study concludes that the development of the AI Ethical Audit Model is a strategic advancement in strengthening ethical governance within Indonesia's digital tax ecosystem, particularly in the operation of the Core Tax Administration System (CTAS). The integration of AI technologies brings significant benefits to administrative efficiency and compliance analytics; however, it also introduces ethical risks that necessitate systematic oversight. By embedding the principles of **fairness, transparency, and accountability**, the proposed model provides a structured mechanism to detect ethical vulnerabilities and enhance the reliability of AI-driven decision-making. The results affirm the urgent need for institutionalized ethical auditing practices as part of digital tax transformation initiatives.

5.2 Theoretical and Practical Implications

Theoretically, this study enriches the scholarly discourse on algorithmic governance by incorporating technological ethical dimensions into the framework of public sector risk-based auditing. The AI Ethical Audit Model extends existing theories on AI ethics by providing a governance-oriented auditing structure specifically designed for government digital systems.

Practically, the model offers clear and applicable guidance for policymakers, especially within the Directorate General of Taxes (DJP). It supports the formulation of ethical AI governance policies, the establishment of an **AI Ethics Oversight Board**, and the implementation of periodic ethical audits. These implications highlight how public institutions can operationalize AI ethics in real-

world digital governance settings to ensure fairness, transparency, and accountability.

5.3 Research Limitations

Several limitations should be acknowledged. First, the study relies on qualitative data obtained from a limited number of expert informants, which may not fully represent the diversity of perspectives across all DJP units or other public sector agencies. Second, the model developed has not yet been empirically tested using quantitative performance indicators, which limits the ability to generalize the model's effectiveness across broader technological environments. Third, the study focuses specifically on the CTAS context, thereby narrowing its applicability unless adapted to other AI-enabled public services. These limitations provide opportunities for further refinement of the model and encourage wider empirical validation.

5.4 Recommendations for Future Research

Future research is encouraged to empirically evaluate the AI Ethical Audit Model using quantitative approaches, such as developing an AI Ethical Compliance Index that measures ethical performance in AI systems across public institutions. Researchers may also expand the model's applicability to other domains, including customs, regional tax administration, public financial management, and digital social services.

Additionally, comparative studies across different countries could provide insights into global best practices in AI ethical auditing. Longitudinal research may further explore how ethical risks evolve over time as AI technologies become more deeply integrated into government operations. By pursuing these directions, future scholarship can contribute to establishing a comprehensive and globally relevant framework for ethical AI governance in the public sector.

Bibliography

Arens, A. A., Elder, R. J., Beasley, M. S., & Hogan,

C. E. (2021). Auditing dan Jasa Assurance (terjemahan). In *Pearson Education Limited*.

Bryson, J. J. , Andriessen, J. E. , & Winfield, A. F. T. (2020). Responsible AI: Implementing ethical principles. *AI & Society*, 35(1), 1–12. <https://doi.org/https://doi.org/10.1007/s00146-019-00879-2>

Cath, C. (2018). *Governing artificial intelligence : ethical , legal and technical opportunities and challenges Subject Areas : Author for correspondence :*

Creswell, J. W.; Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Doshi-Velez, F.; Kim, B. (2018). *Towards a rigorous science of interpretable machine learning*. <https://doi.org/10.48550/arXiv.1702.08608>

Floridi, L., & Cowls, J. (2022). The ethics of artificial intelligence: Principles, challenges, and opportunities. *Philosophy & Technology*, 35(2), 15–31. <https://doi.org/https://doi.org/10.1007/s13347-021-00460-6>

Freeman, R. E. (1984). *Strategic Management: A Stakeholder Approach*. Pitman Publishing.

Jensen, C., & Meckling, H. (1976). *THEORY OF THE FIRM: MANAGERIAL BEHAVIOR , AGENCY COSTS AND OWNERSHIP STRUCTURE I . Introduction and summary In this paper WC draw on recent progress in the theory of (1) property rights , firm . In addition to tying together elements of the theory of e. 3, 305–360.*

Jobin, A., Ienca, M., & Vayena, E. (2019). *Artificial Intelligence : the global landscape of ethics guidelines*.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6), 1–35. <https://doi.org/https://doi.org/10.1145/3457607>

- Nataliawati, R., Yaumi, S., Qosim, Yulianita, F., & Sutiani. (2024). *Implementation of Artificial Neural Networks as a Method for Early Detection of Tax Evasion Behavior in Indonesia*. 4(4), 273–284.
- Neuman, W. L. (2014). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Pearson Education.
- OECD. (2021). *OECD Framework for the Classification of AI Systems*. <https://doi.org/https://doi.org/10.1787/cb6d99c9-en>
- OECD. (2023a). *Tax Administration 2023: Comparative Information on OECD and Other Advanced and Emerging Economies*. <https://doi.org/https://doi.org/10.1787/74d162b6-en>
- OECD. (2023b). *Tax Administration 2023: Comparative Information on OECD and Other Advanced and Emerging Economies*. <https://doi.org/10.1787/74d162b6-en>
- Rahwan, I. (2017). *Society-in-the-Loop : Programming the Algorithmic Social Contract*.
- Sætra, H. S. (2023). AI, ethics and the law: The case for AI accountability. *AI and Ethics*, 3(1), 47–59. <https://doi.org/https://doi.org/10.1007/s43681-021-00109-5>