Article

Human-in-the-Loop (HITL) in AI-Assisted Disclosure

Fatima Salman*

Department of Law, University of Punjab, Jhelum Campus Jhelum, Pakistan
* Correspondence: salmanfatima871@gmail.com

Abstract: Artificial Intelligence (AI) has become an integral part of corporate disclosure processes, changing the way that companies create, validate and communicate information to others. The use of AI in decision making and disclosure processes, has revolutionized the governance, in the legal, financial and public administration sectors. Its critical, especially in high-stakes settings such as financial reporting and legal disclosure requirements. Although AI-assisted disclosure is advantageous in terms of efficiency, scalability and analytical capabilities, it poses governance, auditability and explainability challenges. These dimensions are interdependent and current study explores these dimensions highlighting the need for accountability and compliance of human-in-the-loop (HITL) supervision. The study suggests an integrated architecture which includes governance structures, audit trails, explainability techniques and structured human monitoring. Current research analyzes disclosure processes and introduces a new AI-assisted disclosure approach with a Human-in-the-Loop (HITL) system that allows for human oversight, accountability, and trust. EU AI Act, NIST AI RMF, and emerging compliance and open governance practices, should be reviewed regarding their application to HITL systems to investigate trends and patterns in managing 'black box' risks, audit trails and meaningful explanations. Technical (e.g., XAI techniques such as SHAP/LIME) and procedural safeguards must be combined to provide an effective governance process, holistic HITL supplemented by technical safeguards do not rubber stamp decisions, but instead increase decision quality. Good governance needs to be multi-layered in which there is structured human intervention, regulatory control and technical transparency. A full package approach, combining audit trails, explainability protocols and human review processes is recommended to ensure legally compliant and ethically responsible disclosure when made by AI.

Keywords: AI governance; human-in-the-loop; explainable AI; auditability; regulatory disclosure; accountability

Citation: Fatima Salman. 2026. Human-in-the-Loop (HITL) in AI-Assisted Disclosure. *Digital Social Sciences 3(1),* 1-4. https://doi.org/10.69971/dss.3.1.2026.46

1. Introduction

AI has quickly revolutionized how organizations interact and report communication. The ability of generative AI systems to create annual reports, prepare sustainability disclosures, summarize financial information, respond to regulators and aid management in stakeholder communications. As disclosure becomes more and more reliant on AI, there are opportunities for more efficient disclosure and lower reporting costs. The advent of AI in disclosure is a revolution as companies share important information with regulators, investors and stakeholders. Disclosure involves a wide variety of formal reporting, such as financial reporting, regulatory reporting e.g., securities and exchange commission (SEC) filings, corporate governance statements, sustainability reporting and public government documents. Automating this disclosure process with AI systems raises questions of transparency, accountability, and oversight regarding the risk and compliance verification. As AI systems take on more of the content classification, risk and compliance verification and compliance tasks in these disclosure workflows, new challenges arise in terms of transparency, accountability and oversight.

To aid disclosure processes, improve decision making efficiency and efficiency in disclosure, organizations are implementing machine learning systems, natural language processing (NLP) and generative AI systems (Kirk, Peter and Evan 2025). Although these technologies have significant advantages, they also come with challenges such as transparency, accountability, and legal liability (Ahmed, Seema and Salman 2025). Disclosure activities are the core of accountability process. Disclosures are important for making decisions by investors, regulators, auditors, employees and society. The issue of responsibility for disclosure mistakes, biases, hallucinations, omissions or misstatements arises where the AI systems are involved in the disclosure generation. The difficulty is exacerbated when the reasoning behind the use of a sophisticated AI model can't be easily understood, such as when it is used as a "black box."

Governance is the setting of policies, roles and oversight structures. Auditability provides a way to keep track of every aspect of AI input, process, output, and human involvement. Explainability (via Explainable AI or XAI) allows to understand the reasoning of AI with stakeholders. Human-in-the-Loop (HITL) involves human decision-making in key stages of the process, ensuring that automation can only be a supporting role in high-stakes contexts. AI governance frameworks focus on transparency, accountability, explainability, and oversight by human beings as core principles to ensure trustworthy AI. Explainability, accountability, and transparency are all key properties of trustworthy AI systems that are included in the NIST AI Risk Management Framework. Transparency and accountability are often mentioned in organizational frameworks.

The use of AI in disclosure systems raises questions about traditional governance structures as the production of information is shared, spread, and distributed between the developer, the system manager, the auditor, and the AI system (Floridi et al. 2018). Therefore, scholars have suggested that AI governance should not just focus on technical regulation, but also include organizational responsibility and regulation (Mittelstadt et al. 2016).

The human-in-the-loop review process becomes a key continuity process, whereby the human is integrated into the process with their knowledge, expertise and experience to enhance all three aspects. The EU AI Act (2024) and NIST AI Risk Management Framework (2023), pose requirements for human oversight and transparency of documentation of high-risk AI systems. But there are sizeable disconnects between the regulations and the actual practice in terms of the amount and level of human oversight and the amount of "meaningful" human oversight versus "theater" of the rules.

Current study examines the role of governance to ensure that disclosure using AI is auditable, explainable, accountable and with human oversight. It suggests a governance system comprising a mix of technical controls, organizational processes and governance mechanisms, that help promote responsible deployment.

2. Theoretical Foundations

The AI governance concepts and theories are based on agency theory, stakeholder theory, legitimacy theory, information asymmetry, and algorithmic accountability. Socio-technical governance models acknowledge interactions between technology, institutions and people.

Agency theory implies that disclosure mechanisms minimize information asymmetries between management and stakeholders. AI systems can have a positive impact on the quality of disclosures, but can also create hidden decision-making processes. Trust and transparency are key attributes of the stakeholder theory. Information asymmetry between managers and stakeholders will be diminished by disclosure (Jensen and Meckling 1976). AI-assisted disclosure can enhance information flows, due to its potential to provide better accuracy and timeliness of information, but it can also generate new asymmetries, as the algorithms used are opaque (Busuioc 2021).

Stakeholder theory highlights that organizations need to build trust with various stakeholder groups by communicating in a transparent manner (Freeman 2015). Explainability is thus crucial as stakeholders have come to expect that organizations must provide explanations for decisions made using AI (Karanxha and Paulinus 2025).

Legitimacy theory focuses on the society's expectations on ethical use of AI. According to this theory, the organizations try to earn the acceptance of the society by adopting the norms and expectations of the society (Suchman 1995). Organizations could be expected to show ethical governance and responsible use of automated technologies with the increasing use of AI systems (Jobin, Marcello and Effy 2019).

AI-assisted disclosure is helpful being faster processing, scalability, consistency, predictive analytics, and better data processing. However, pose challenges of algorithmic bias, hallucinations, data quality concerns, cyber security concerns, model drift, and accountability concerns. There is need, therefore, to balance between innovation and governance safeguards in organizations. Validation, monitoring and independent review are needed for effective disclosure systems.

3. Governance of AI-Assisted Disclosure

AI governance is the systems, policies, and processes that govern the design, deployment, and monitoring of AI systems (Databricks 2026). Generally, good governance systems include elements of transparency, accountability, fairness and human oversight (OECD 2019). Lifecycle management strategies that include risk assessment, involve stakeholders, and involve ongoing monitoring are highlighted across the entire lifespan of AI deployment (Herrera-Poyatos et al. 2026). These frameworks acknowledge that governance is something that happens all of the time and not just one-time compliance.

The role of the Boards of Directors in the supervision of algorithmic systems and AI-assisted disclosures becomes more important as the Boards are increasingly held accountable for each of these aspects (Ahmed, Seema and Salman 2025). This is a step toward a new frontier of corporate governance accountability-algorithms. Risk assessment, transparency requirements, ethics guidelines and compliance measures should be included in governance arrangements.AI disclosure systems can create narrative disclosures, condense financial statements, detect irregularities, and help with compliance reporting (Kirk, Peter and Evan 2025). These features facilitate companies to deal with extensive amounts of information more efficiently than conventional manual methods.

However, there are a various challenge. Algorithmic bias can result from training data that reflect inequities in the past, producing skewed results (Barocas and Andrew 2016). In addition, LLMs can generate hallucinated or inaccurate information, posing substantial disclosure risks (OpenAI 2023).

The ability to review and validate the outputs and systems of AI independently is known as auditability (Verma, Kirtan and Eva 2025). Comprehensive documentation, traceability and reproducibility mechanisms are critical for effective auditability. An audit trail is especially significant as it establishes chronological information of inputs, outputs, and system decision making (Raji, Andrew and Rebecca et al. 2020). These records allow for the review of governance needs and to identify compliance pathways. Lifecycle auditing should encompass data collection, model development, deployment, monitoring and retirement. AI systems are not yet capable of replacing human expertise completely in a professional judgment context for complex situations within an organization (Golubovskij 2024).

Explainable AI is a method to make the decisions made by algorithms understandable to human users. SHAP, LIME, feature importance analysis, and counterfactual explanations are methods that facilitate understanding of how outputs are produced by stakeholders. Explainability helps to build trust, accountability, legal defensibility, and regulatory compliance. The goal of XAI is to provide humans with an understanding of algorithmic decisions (Arrieta, Natalia and Javier et al. 2020). The issue of explainability is of specific relevance in disclosure settings, since, in such cases, it is important to understand the procedures used to derive and disclose material information. Several techniques have been developed to enhance transparency, including SHAP, LIME and counterfactual explanations (Molnar 2024). The techniques are used to help users understand what factors are affecting the outputs from the AI and evaluate the reliability of the system's recommendations. Empirical studies have shown that explainable systems boost user trust and aid in more effective auditing (Anomah 2026). In many cases, though, there are compromises between model complexity and interpretability (Rudin 2019).

4. Human-in-the-Loop Review

Human-in-the-loop (HITL) systems involve (Lazaros, Aristidis and Sotiris et al. 2026) human involvement in the decision-making process of AI systems. Where errors have legal or financial implications, such as in high risk disclosure situations, HITL approaches are especially significant. Human reviewers can detect contextual problems, ethical issues, disclosure risks, and other problems that automated systems might not detect. Human oversight requires a form of control over and a capacity to modify or deny the outputs of these AI systems. Human intervention greatly enhances the accountability and minimizes the risks of automated decision-making (Amershi et al. 2014). This is more than a "symbolic review." Human reviewers must have the authority to challenge, edit or reject the output of the AI. There are also documentation and escalation procedures that must be put in place to enable effective monitoring.

The role of the Boards of Directors in the supervision of algorithmic systems and AI-assisted disclosures becomes more important as the Boards are increasingly held accountable for each of these aspects (Ahmed, Seema and Salman 2025). This is a step toward a new frontier of corporate governance accountability-algorithms. Risk assessment, transparency requirements, ethics guidelines and compliance measures should be included in governance arrangements. AI disclosure systems can create narrative disclosures, condense financial statements, detect irregularities, and help with compliance reporting (Kirk, Peter and Evan 2025). These features facilitate companies to deal with extensive amounts of information more efficiently than conventional manual methods.

However, there are a various challenge. Algorithmic bias can result from training data that reflect inequities in the past, producing skewed results (Barocas and Andrew 2016). In addition, LLMs can generate hallucinated or inaccurate information, posing substantial disclosure risks (OpenAI 2023).

The ability to review and validate the outputs and systems of AI independently is known as auditability (Verma, Kirtan and Eva 2025). Comprehensive documentation, traceability and reproducibility mechanisms are critical for effective auditability. An audit trail is especially significant as it establishes chronological information of inputs, outputs, and system decision making (Raji, Andrew, and Rebecca et al. 2020). These records allow for the review of governance needs and to identify compliance pathways. Lifecycle auditing should encompass data collection, model development, deployment, monitoring and retirement. AI systems are not yet capable of replacing human expertise completely in a professional judgment context for complex situations within an organization (Golubovskij 2024).

Explainable AI is a method to make the decisions made by algorithms understandable to human users. SHAP, LIME, feature importance analysis, and counterfactual explanations are methods that facilitate understanding of how outputs are produced by stakeholders. Explainability helps to build trust, accountability, legal defensibility, and regulatory compliance. The goal of XAI is to provide humans with an understanding of algorithmic decisions (Arrieta, Natalia and Javier et al. 2020). The issue of explainability is of specific relevance in disclosure settings, since, in such cases, it is important to understand the procedures used to derive and disclose material information. Several techniques have been developed to enhance transparency, including SHAP, LIME and counterfactual explanations (Molnar 2024). The techniques are used to help users understand what factors are affecting the outputs from the AI and evaluate the reliability of the system's recommendations. Empirical studies have shown that explainable systems boost user trust and aid in more effective auditing (Anomah 2026). In many cases, though, there are compromises between model complexity and interpretability (Rudin 2019).

5. Integrated Governance

A combination of governance structures, auditing procedures, explanations and human involvement. The model emphasizes forward and backward loops, and accountability of the organization, in the AI lifecycle. The regulators need to formulate common rules for the use of AI, business actors should improve their governance mechanisms, and auditors need to enhance their technical expertise in AI assurance. A co-operative professional education and collaboration is key. The importance of transparency, explainability, and human oversight in AI governance is increasingly gaining traction. Harmonized standards can minimize compliance uncertainty and the level of conformity between jurisdictions. Organizations need to build their governance and AI literacy programs, and invest in their workforce skills, to make sure that AI is deployed responsibly (OECD 2019). Technical skills are also essential for auditors to assess increasingly complex AI systems (Verma, Kirtan and Eva 2025). Explainability indicators, auditing strategies, governance models of maturity and cross-jurisdictional approaches to regulation should be explored as topics for future study. Comparative study can be used to further shed light on best practices.

6. Conclusion

AI-assisted disclosure has both governance challenges and opportunities. To be trusted, deployment needs to be integrated with governance, auditability, explainability and human oversight. There is a strong case for hybrid human-AI governance systems to offer the best pathway towards accountable and transparent disclosure practices. AI-powered disclosure is a major leap forward for organizations in their reporting abilities. The incorporation of AI into disclosure processes, however, poses significant challenges of accountability, transparency, auditability and explainability. Governance mechanisms are crucial to ensure that AI systems build, not destroy, stakeholders' trust. There is a need for a mix of explainability mechanisms, wide-spread audit trails and human-in-the-loop review mechanisms to achieve effective governance. There is a need to keep human oversight at the heart of disclosure decision making, and have AI to be an assistive tool, not the decision maker. Businesses that effectively in corporate governance, auditable, and explainable elements into their disclosure practices using AI will be more likely to achieve regulatory compliance, stakeholder trust, and organizational legitimacy.

References

Ahmed, Fazail Asrar, Seema Gul, and Salman Shahzad. 2025. Ensuring Accountability and Transparency in AI-Driven Corporate Governance. *International Journal of Social Sciences Bulletin* 3:330-341. https://doi.org/10.5281/zenodo.15378921

Amershi, Saleema, Maya Cakmak, W. Bradley Knox, and Todd Kulesza. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35 (4): 105–120. https://doi.org/10.1609/aimag.v35i4.2513

Anomah, Sampson. 2026. Bridging the Interpretability Gap: Integrating Explainable AI into Public Sector Audit Systems in Emerging Economies. *Accounting Research Journal* 39: 1-27. https://doi.org/10.1108/ARJ-08-2025-0302

Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez and Javier Del Ser, drien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion* 58: 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Barocas, Solon, and Andrew D. Selbst. 2016. Big Data's Disparate Impact. *California Law Review* 104: 671–732. https://doi.org/10.15779/Z38BG31

Busuioc, Madalina. 2021. Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review* 81: 825–836. https://doi.org/10.1111/puar.13293

Databricks. 2026. AI Governance Best Practices: Responsible and Effective AI Programs. *Databricks*. https://www.databricks.com/blog/ai-governance-best-practices-how-build-responsible-and-effective-ai-programs

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendation. *Minds and Machines* 28: 689–707. https://doi.org/10.1007/s11023-018-9482-5

Freeman, R. Edward. 2015. *Strategic Management: A Stakeholder Approach*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139192675

Golubovskij, Vladislav. 2024. Explainability and Transparency of AI in Auditing. Bachelor's thesis, University of Twente. https://essay.utwente.nl/106877/

Herrera-Poyatos, Andrés, Javier Del Ser, Marcos López de Prado, and Jesús L. Lobo. 2026. A Framework for Responsible AI Systems. *Arxiv*. https://arxiv.org/abs/2503.04739

Jensen, Michael C., and William H. Meckling. 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics* 3: 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence* 1: 389–399. https://doi.org/10.1038/s42256-019-0088-2

Karanxha, Giulia, and Paulinus Ofem. 2025. Evaluating Transparency in AI Systems. *IEEE International Conference Proceedings*. IEEE. https://doi.org/10.1109/ICAI60593.2025.11330146

Kirk, Marcus, Peter Molk, and Evan Pondel. 2025. AI in Corporate Disclosure. *University of Florida Levin College of Law Research Paper* 26. https://doi.org/10.2308/HORIZONS-2024-074

Lazaros, Konstantinos, Aristidis G. Vrahatis and Sotiris Kotsiantis. 2026. Human-in-the-Loop Artificial Intelligence: A Systematic Review of Concepts, Methods, and Applications. *Entropy* 28 (4): 377. https://doi.org/10.3390/e28040377

Mittelstadt, Brent Danial, Patrick Allo, and Luciano Floridi. 2016. The Ethics of Algorithms: Mapping the Debate. *Big Data & Society* 3: 1–21. https://doi.org/10.1177/2053951716679679

Molnar, Christoph. 2024. *Interpretable Machine Learning*. 3rd Ed. https://leanpub.com/interpretable-machine-learning

OECD. 2019. OECD AI Principles overview. https://oecd.ai/en/ai-principles

OpenAI. 2023. GPT-4 Technical Report. https://cdn.openai.com› papers

Raji, Inioluwa Deborah, Andrew Smart, and Rebecca N. White et al. 2020. Closing the AI Accountability Gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the ACM Digital Library*. https://dl.acm.org/doi/10.1145/3351095.3372873

Rudin, Cynthia. 2019. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence* 1: 206–215. https://doi.org/10.1038/s42256-019-0048-x

Suchman, Mark C. 1995. Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review* 20: 571–610. https://doi.org/10.5465/AMR.1995.9508080331

Verma, Himanshu, Kirtan Padh, and Eva Thelisson. 2025. Can AI Be Auditable? *Arxiv*. https://arxiv.org/abs/2509.00575