Human-in-the-Loop AI: Rethinking Automation Ethics in Decision-Sensitive Domains Case Study of the Education, IT and Non-for-Profit sectors.

Nankyer Sarah Joseph, Mohammed Nasiru Yakubu

Department of Information Systems, School of IT and Computing, American University of Nigeria, Yola Mohammed Nasiru Yakubu Arden University Middlemarch Business Park, Coventry. CV3 4FJ

DOI: https://doi.org/10.51584/IJRIAS.2025.1010000027

Received: 23 September 2025; Accepted: 30 September 2025; Published: 30 October 2025

ABSTRACT

This study develops and applies the Human-in-the-Loop (HITL) Ethical Assessment Framework (EHAF) to examine the ethical sufficiency of HITL artificial intelligence (AI) across education, information technology (IT), and non-profit sectors. The research objective was to evaluate how effectively HITL practices safeguard human values in decision-sensitive contexts and to identify sector-specific challenges that may compromise ethical adequacy. Adopting a qualitative thematic approach, we analyzed survey responses from professionals in the three sectors. Responses were coded against the four diagnostic dimensions of EHAF Impact Severity, Contextual Ambiguity, Human Agency, and Transparency & Auditing while also allowing for the identification of emergent themes. Retroductive reasoning was used to move beyond surface patterns to uncover generative mechanisms shaping HITL practices. Findings demonstrate sectoral variation in how HITL systems are operationalized and valued. In education, ethical sufficiency is closely tied to human oversight given the high stakes of student outcomes and the importance of cultural contextualization. In the non-profit sector, transparency and auditing dominate due to donor accountability pressures and reporting requirements. IT organizations, by contrast, privilege efficiency and scalability, but often provide weaker safeguards for human agency and oversight. Across all sectors, emergent themes such as training, trust, infrastructure readiness, and donor influence were found to condition HITL adequacy. Generative mechanisms identified include institutional role ambiguity, donor pressure, cultural misalignment, and capacity constraints. The study concludes by proposing an extension to EHAF that incorporates a fifth dimension, Capacity and Governance Context to better capture systemic and institutional influences. Conceptually, the paper refines the assessment of HITL ethics, while practically offering sector-specific recommendations to strengthen oversight, accountability, and trust in AI-enabled decision-making.

Keywords: Human-in-the-Loop, AI Ethics, Automation, Decision-Sensitive Domains, Responsible AI,

INTRODUCTION

Artificial intelligence (AI) is increasingly shaping decision-making across critical domains such as education, information technology, and the non-profit sector. While automation promises efficiency and scalability, the ethical risks of delegating sensitive decisions entirely to algorithms have become more apparent [1]. Concerns around bias, opacity, and accountability gaps are particularly acute in contexts where outcomes directly affect human welfare and equity [2] These risks highlight the pressing need to balance automation with human oversight.

Within the information systems (IS) discipline, scholars emphasize that AI systems are sociotechnical rather than purely technical artifacts, and thus require governance frameworks that account for both human and machine roles in decision-making [3]. One promising approach is Human-in-the-Loop (HITL), where human judgment is embedded within algorithmic processes to provide contextual reasoning, safeguard against harmful outcomes, and reinforce accountability (Rahwan et al., 2019).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue X October 2025

However, current AI governance models often adopt a techno-centric perspective that overlooks organizational, cultural, and ethical complexities [4]. While automation is framed as a means of minimizing human error, little attention has been paid to how HITL practices can reshape ethical responsibility and trust in decision-sensitive contexts. Moreover, there is limited qualitative empirical research exploring HITL implementation in diverse organizational domains, particularly within education, IT services, and the non-profit sector, where decision stakes are high but resources and institutional safeguards may vary [5].

This paper addresses these gaps by examining how HITL approaches can enable more responsible AI in decision-sensitive domains. Drawing on a qualitative case study design, with data collected from organizations in IT, education, and the non-profit sector, we investigate how human oversight interacts with automated decision-making in practice. The study seeks to answer the guiding question:

How can Human-in-the-Loop AI practices enhance ethical responsibility and trust in decision-sensitive domains?

The contribution of this paper is threefold. First, it extends IS research on sociotechnical governance of AI by theorizing HITL as an ethical safeguard. Second, it offers empirical insights into the opportunities and challenges of HITL implementation across multiple domains. Finally, it provides policy and organizational recommendations for embedding human oversight into AI systems, thereby contributing to ongoing debates on responsible AI governance.

LITERATURE REVIEW AND THEORETICAL FRAMING

The integration of Artificial Intelligence (AI) in decision-sensitive fields such as information technology (IT), education, and the non-profit sector has garnered increasing attention, primarily concerning the ethical implications of Human-in-the-Loop (HITL) systems. HITL approaches, which incorporate human judgment into AI decision-making processes, are pivotal in ensuring that AI systems operate within ethical boundaries and align with societal values.

Benefits of HITL Systems

In the education sector, HITL systems facilitate personalized learning experiences by adapting to individual student needs and providing real-time feedback. This adaptability enhances student engagement and supports diverse learning styles [6]. Similarly, in the non-profit sector, AI tools enable organizations to streamline operations, improve donor engagement, and optimize resource allocation, thereby increasing operational efficiency.

Furthermore, HITL systems contribute to the ethical deployment of AI by allowing human oversight to correct biases, ensure fairness, and maintain accountability in decision-making processes. This human oversight is crucial in high-stakes applications where AI decisions can significantly impact individuals' lives.

Challenges Associated with HITL Systems

Despite their advantages, HITL systems present several challenges. In education, the reliance on AI tools may lead to concerns about data privacy, algorithmic bias, and the potential erosion of teacher-student relationships [7]. Additionally, the effectiveness of HITL systems is contingent upon the quality of human input; inadequate or biased human judgment can perpetuate existing inequalities in educational outcomes.

In the non-profit sector, the adoption of AI technologies can be hindered by limited financial resources, lack of technical expertise, and resistance to change within organizations. Moreover, the ethical implications of using AI in non-profit settings, such as the potential for exploitation of vulnerable populations and the need for transparency, require careful consideration.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue X October 2025

Applications of HITL Systems in Education and Non-Profit Sectors

In education, HITL systems are applied in adaptive learning platforms, automated grading systems, and virtual teaching assistants. These applications aim to enhance learning outcomes by providing personalized support and reducing administrative burdens on educators [6].

In the non-profit sector, HITL systems are utilized in areas such as donor segmentation, campaign optimization, and impact assessment. By integrating human expertise with AI capabilities, organizations can better understand donor behavior, tailor communications, and measure the effectiveness of their initiatives.

Ethical Dilemmas in HITL Systems

The deployment of HITL systems raises several ethical dilemmas. In education, issues related to data privacy, consent, and the potential for algorithmic bias in student assessments are of paramount concern [7]. The use of AI tools must be transparent, with clear guidelines on data usage and mechanisms for accountability.

In the non-profit sector, ethical considerations include the equitable access to AI technologies, the potential for reinforcing existing power imbalances, and the need for inclusive decision-making processes. Organizations must ensure that AI applications align with their mission and values, promoting social good without exploiting vulnerable populations.

The integration of HITL systems in education and the non-profit sector offers significant benefits, including personalized learning experiences and enhanced operational efficiency. However, these advantages must be weighed against the associated challenges and ethical dilemmas. A balanced approach, incorporating human oversight and ethical considerations, is essential for the responsible deployment of AI technologies in these sectors.

Human-in-the-Loop (HITL): definitions & modalities

HITL refers to system architectures that deliberately integrate human judgment/expertise into the lifecycle of an AI system at critical decision points:

(a) **pre-decision** (data/model design and feature selection),

(b) **real-time intervention** (human overrides /interventions during automated operation), and

(c) **post-decision oversight** (audits, appeals, human review).

HITL spans a spectrum from heavy human control to lightweight human validation; the ethical promise derives from preserving contextual reasoning and accountability that pure automation lacks [8].

Automation bias and oversight failure

A well-documented risk in HITL contexts is automation bias decision makers tend to over-rely on automated outputs, reducing vigilance and failing to catch algorithmic errors. Systematic reviews demonstrate automation bias across domains (healthcare, aviation, administrative systems) and identify mediators (display design, workload, expertise) and mitigators (confidence displays, training). This means that mere human presence is not sufficient; humans must be empowered and equipped to challenge machine outputs [9]

Human–ML augmentation & IS perspectives

Recent IS scholarship argues for nuanced typologies of human–ML augmentation (e.g., reactive oversight, proactive oversight, informed reliance, supervised reliance) [10]

and calls for IS research to rethink classic assumptions in light of ML's unique properties (data-trained models, non-deterministic behaviors). Studies have shown how automation fairness failures invite managerial and design strategies that center human–ML collaboration rather than binary automation/no-automation choices [11].

Theoretical lens: sociotechnical systems + normative governance

We frame HITL as a sociotechnical governance mechanism: ethical outcomes arise from the interplay of algorithmic affordances, human judgment, organizational structures, and regulatory context. This aligns with IS sensibilities that emphasize technology-organization-environment interactions and the need for institutional scaffolding to secure accountable outcomes [12] The E U AI Act's explicit human oversight requirement further signals policy momentum for operationalizing HITL in high-risk systems.

RESEARCH DESIGN & METHODOLOGY

Research strategy multiple qualitative case study

This study adopts a qualitative multiple case study [12] to compare how HITL is designed, enacted, and sustained across three decision-sensitive domains: Education, IT, and Non-Profit. This design allows theory-building about conditions for HITL sufficiency and identification of demi-regularities across contexts while preserving attention to domain peculiarities.

Case selection & sampling

Purposive sampling of six organizations within the IT, Education and Non-for Profits sectors in Nigeria, (two per domain) selected for active use of ADS with some human oversight mechanism (formal or informal).

Within each case, we interview 2 participants across roles: designers/engineers, administrators/managers, frontline human reviewers, and affected stakeholders (teachers, beneficiaries). Total planned interviews: 6. Documents (policies, decision logs, audit reports) and limited observations were reviewed.

Data collection instruments and Analysis

Semi-structured, open-ended questionnaires were administered to experts in the IT, Education, and Non-profit sectors. The instrument was designed to capture perceptions of AI oversight, ethical concerns, and practical experiences with Human-in-the-Loop (HITL) systems. Questions explored decision-making workflows, accountability practices, and contextual challenges. Responses were documented, compiled, and later transcribed into structured datasets for analysis. Collected responses from the spreadsheet was loaded into NVivo for categorization into themes and final analysis [13].

Analytical lens: Ethical HITL Assessment Framework (EHAF) proposed

As Artificial Intelligence (AI) systems become increasingly integrated into critical sectors such as healthcare, finance, and education, ensuring ethical decision-making is paramount [14]. Human-in-the-Loop (HITL) systems have emerged as a key approach for embedding human judgment, oversight, and intervention into AI workflows, helping to mitigate bias and prevent harmful outcomes [15].These systems are particularly important in domains where decisions carry significant ethical implications, or where AI alone cannot fully interpret complex, context-sensitive information [16].

To systematically evaluate the ethical adequacy of HITL systems, we propose the **Ethical Human-in-the-Loop Assessment Framework (EHAF)**, comprising four diagnostic dimensions:

Impact Severity – This dimension assesses the stakes of AI-influenced decisions, emphasizing the need for human oversight in high-consequence scenarios. HITL systems are especially critical in domains such as healthcare and student evaluation, where errors may have significant consequences [17].

Contextual Ambiguity – This dimension considers whether AI operates in environments requiring nuanced human interpretation. AI systems often struggle with contextually ambiguous scenarios, making human intervention essential for ethical decision-making [18].

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue X October 2025

Human Agency – This dimension evaluates whether human reviewers are empowered, trained, and authorized to override AI decisions. Evidence suggests that ethical outcomes improve when humans are given meaningful decision-making authority [19].

Transparency & Auditing – This dimension examines whether AI decisions are traceable and auditable. HITL mechanisms support accountability and facilitate post hoc review of AI decisions, which is crucial for compliance and ethical assurance (Rahwan, 2018).

The EHAF was employed as an analytic coding frame during cross-case analysis to determine HITL adequacy and to derive policy and design recommendations [20]

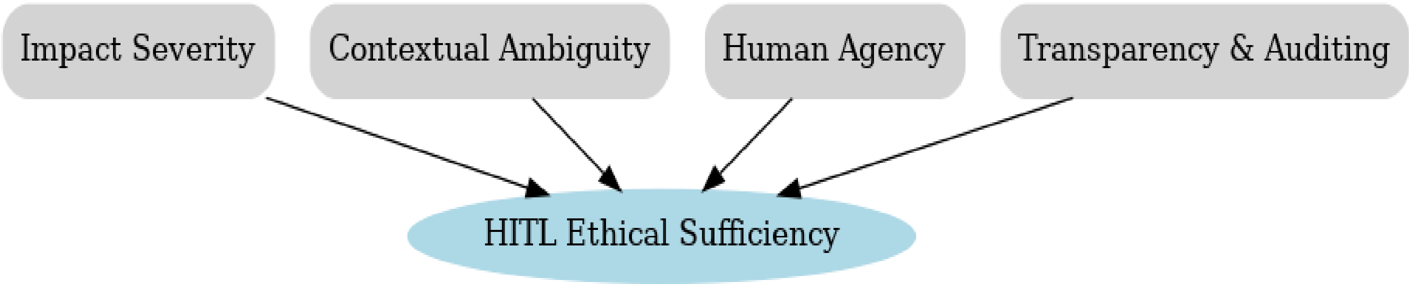

Figure 1:Ethical HITL Assessment Framework (EHAF) Source: Authors work.

RESULTS

Thematic analysis of the survey responses generated insights into how Human-in-the-Loop (HITL) AI is perceived and operationalized across the Education, IT, and Non-profit sectors. Using a hybrid deductive–inductive approach, we mapped responses to the four components of the HITL Ethical Assessment Framework (Impact Severity, Contextual Ambiguity, Human Agency, and Transparency & Auditing) while also identifying emergent inductive themes such as Training & Capacity, Trust, and Infrastructure & Cost.

Sector-Level HITL Ethical Sufficiency

Sector	Impact Severity	Contextual Ambiguity	Human Agency	Transparency & Auditing	Composite %	Interpretation
Education	High	Moderate	Moderate	Moderate	~70%	Moderate–Strong
IT	Low	Low	Weak	Moderate	~45%	Weak
Non-profit	Moderate	Moderate	Moderate	Strong	~60%	Moderate

Table 1: Sector-Level HITL Scores: (Derived from coded scores: see sector_scores.xlsx)

Table 1 summarizes sector-level scores derived from coded responses. Education exhibited the highest sufficiency (≈70%), reflecting strong recognition of decision sensitivity and greater emphasis on human oversight. Non-profit organizations showed moderate sufficiency (≈60%), driven by donor accountability mechanisms but constrained by infrastructural challenges. IT respondents reported the lowest sufficiency (≈45%), indicating weaker integration of human oversight and contextual adaptation.

Table1. Sector-Level HITL Scores (see above table)

The sectoral contrasts are further visualized in Figure 1, which illustrates the distribution of component-level scores. Education scored strongest on Impact Severity, while Non-profits were strongest on Transparency & Auditing. The IT sector consistently underperformed across components, especially in Human Agency.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue X October 2025

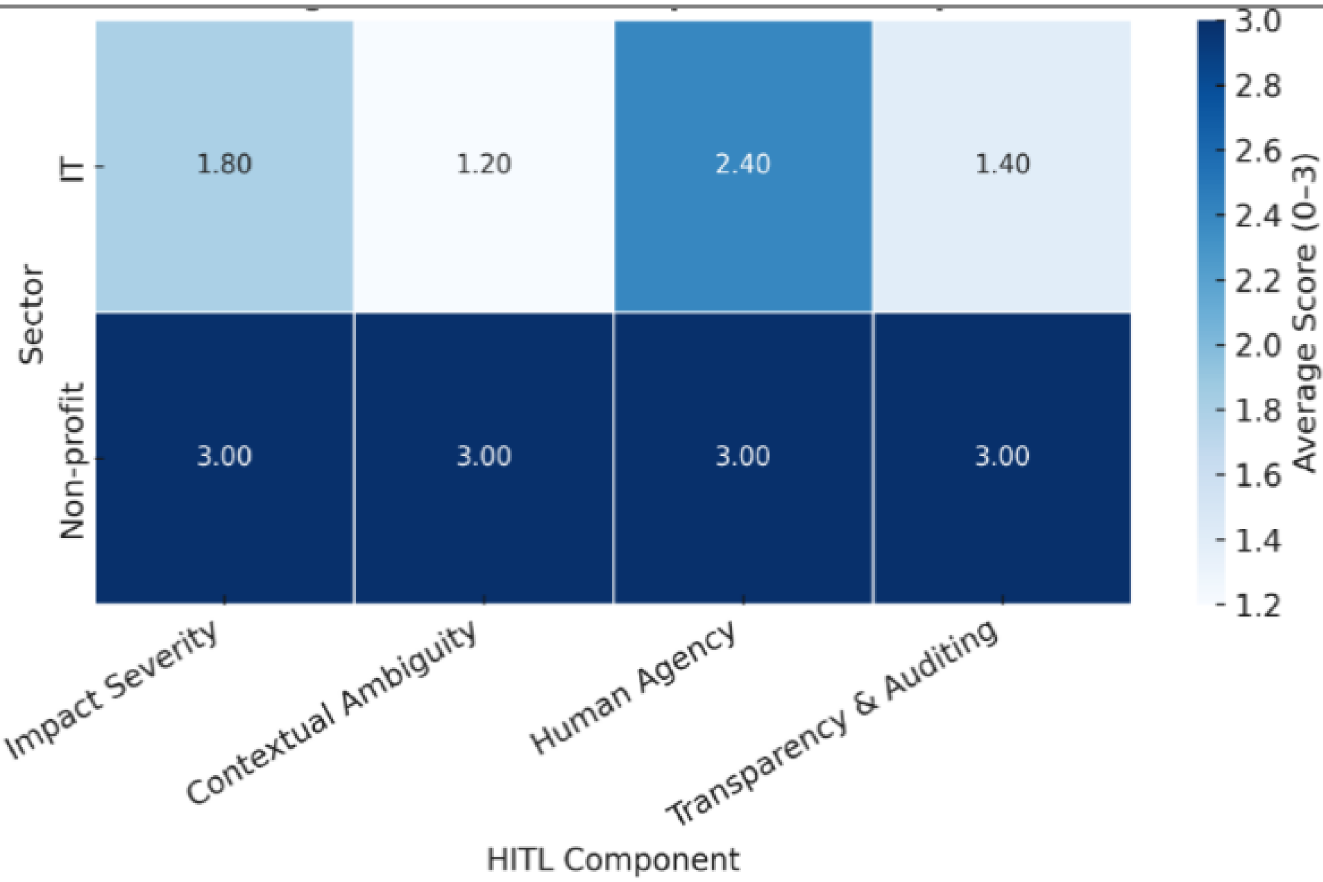

Figure 2:Sector × Component Heatmap (see above chart)

Deductive Findings: HITL Framework Components Impact Severity.

Education respondents emphasized the irreversible consequences of AI-driven decisions on students: "If the algorithm sends a child home, that can change their life we cannot let a machine make that decision alone." Non-profits echoed this awareness in relation to community programs, whereas IT respondents often described system errors as "minor and recoverable." This reflects a divergence in how sectors perceive risk and consequence.

Contextual Ambiguity.

Ambiguity appeared across all sectors but manifested differently. In Education, it reflected cultural nuances "The AI didn't recognize the cultural nuance what looks like absenteeism was actually a festival." In Non-profits, ambiguity was tied to donor requirements conflicting with local needs, while IT respondents referred to technical edge cases.

Human Agency.

The strongest evidence of human oversight came from Education and Non-profits, where teachers and program officers retained override authority: "Program officers override AI outputs if they conflict with donor accountability rules." By contrast, IT respondents often expressed reliance on automated recommendations: "We mostly accept what the system suggests."

Transparency & Auditing.

Transparency was highly salient for Non-profits, where audit trails were maintained for donor accountability: "We maintain audit trails for donors." Education institutions relied on dual logging practices, while IT respondents pointed to opaque audit logs "Logs are there, but no one outside IT understands them."

These findings are summarized in Table 2, which presents sectoral exemplars for each framework component.

Table 2:Exemplar Quotes by Framework Component (Full exemplar dataset: exemplar_sentences_by_code_and_sector.xlsx)

Component	Education Example	IT Example	Non-profit Example
Impact Severity	"If the algorithm sends a child home, that can change their life we cannot let a machine make that decision alone."	"Most system errors are minor and recoverable."	"Decisions affect community access, so we treat them with caution."
Contextual Ambiguity	"The AI didn't recognize the cultural nuance what looks like absenteeism was actually a festival."	"System doesn't account for unusual local inputs."	"Donor requirements often clash with local realities."
Human Agency	"Teachers must always sign off before final decisions."	"We mostly accept what the system suggests."	"Program officers override AI outputs if they conflict with donor accountability rules."
Transparency & Auditing	"We keep manual logs alongside AI records."	"Logs are there, but no one outside IT understands them."	"We maintain audit trails for donors."

Table 2. Exemplar Quotes by Framework Component (see above table)

Inductive Findings: Emergent Themes

Beyond the deductive framework, inductive coding revealed themes that expand the ethical discussion of HITL systems:

Training & Capacity: Respondents across all sectors highlighted skill gaps in interpreting AI outputs, with Education respondents especially emphasizing staff training.

Trust: Trust in AI recommendations was uneven, with some respondents expressing confidence while others described skepticism or discomfort.

Infrastructure & Cost: IT and Non-profit respondents frequently pointed to connectivity, affordability, and sustainability challenges.

Policy & Governance: Non-profit respondents stressed the need for ethical policies and compliance mechanisms.

Donor/Stakeholder Influence: Donor expectations strongly shaped HITL implementation in Non-profits, often reinforcing accountability but also constraining flexibility.

Table 3:Top Emergent Inductive Themes: (Counts from code_counts_overall.xlsx)

Theme	Description	Frequency (mentions)	Sector Presence
Training & Capacity	Calls for more staff training to handle HITL systems	High	All, esp. Education
Trust	Mixed trust/distrust in AI recommendations	High	All sectors
Infrastructure & Cost	Connectivity, affordability, system sustainability	Moderate	IT & Non-profit
Policy & Governance	Formal policies, guidelines, ethics frameworks	Moderate	Non-profit heavy
Donor/Stakeholder Influence	Donor requirements shape HITL use	Moderate	Non-profit

Table 3 presents these emergent themes with frequency counts and sector presence.

Table 3. Top Emergent Inductive Themes (see above table)

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue X October 2025

Overall Patterns

Figure 3 highlights the most frequently occurring codes across the dataset, underscoring the centrality of Human Agency, Transparency, Training, and Trust in shaping perceptions of HITL sufficiency.

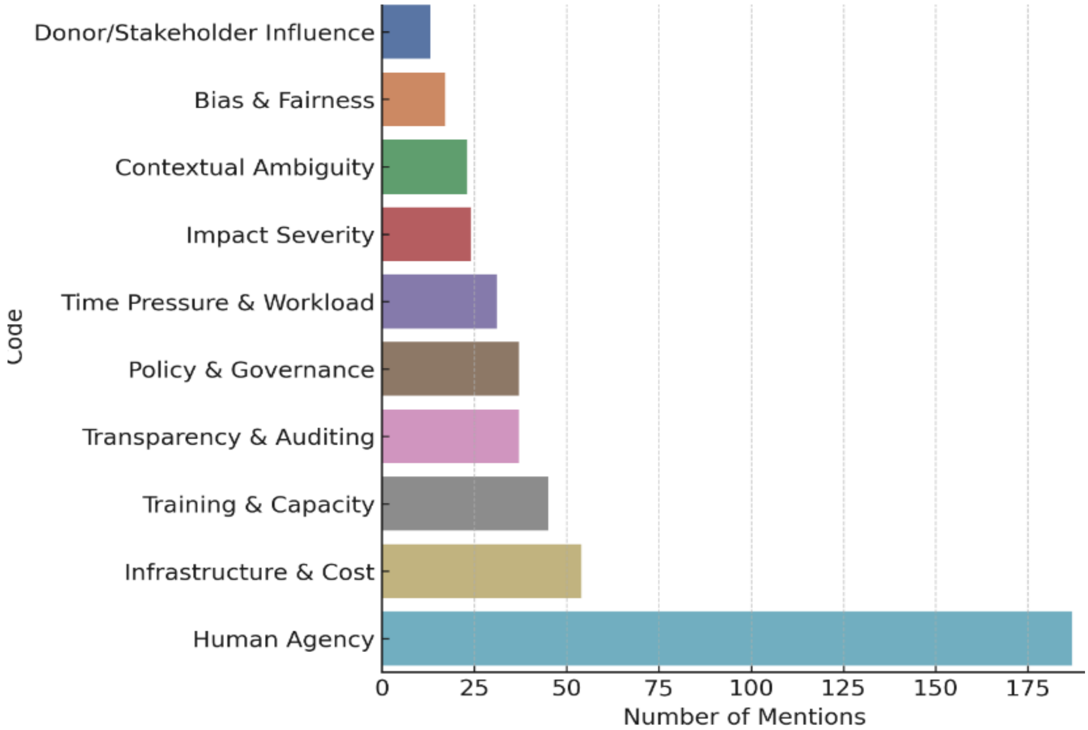

Figure 3:Top 10 Codes by Total Mentions (see above chart)

Taken together, the results reveal that while HITL principles are acknowledged across all three sectors, their operationalization is uneven. Education emphasizes the gravity of decision-making, Non-profits emphasize accountability structures, while IT emphasizes efficiency, often at the expense of human agency.

DISCUSSION

Summary of key findings

This study set out to examine how Human-in-the-Loop (HITL) AI is perceived and operationalized across Education, IT, and Non-profit sectors, and to test the analytic utility of the HITL Ethical Assessment Framework (Impact Severity; Contextual Ambiguity; Human Agency; Transparency & Auditing). The results show consistent recognition of HITL principles across sectors, but reveal important differences in how those principles are enacted. Education respondents showed the strongest concern for impact severity and clearer human-override practices; Non-profit respondents emphasized transparency and audit trails driven by donor accountability; IT respondents prioritized efficiency and technical transparency, but routinely showed weaker human agency and contextual adaptation. Emergent themes especially training and capacity, trust, infrastructure and cost, and vendor dependence cut across sectors and colored how HITL sufficiency was experienced in practice.

Below, we interpret these patterns retroductively, proposing the deeper mechanisms that plausibly generate the observed sectoral differences. We then discuss how these mechanisms confirm, complicate, and extend the HITL framework, and outline practical and research implications.

Retroductive explanation: generative mechanisms behind sectoral differences

Institutional role ambiguity in IT → weak human agency

Pattern observed: IT respondents frequently reported automated workflows and acceptance of system outputs; explicit human override or human review procedures were less common.

Retroductive mechanism: In many IT organizations, governance privileges technical uptime, automation, and rapid response. Organizational roles emphasize system maintenance and scalability rather than deliberative decision-making. This institutional orientation creates role ambiguity for non-technical staff: either no clear decision owner exists, or decision authority defaults to technical teams whose priorities favor performance metrics (uptime, throughput) over deliberative oversight. When role boundaries are unclear and incentives favor automation, human actors are less empowered to intervene. This reduces real human agency even where technologies are nominally "human-in-the-loop."

Evidence link: IT respondents' exemplar statements about accepting system recommendations and difficulty interpreting logs suggest humans are present but lack authority or context to act. The mechanism explains low Human Agency scores despite moderate Transparency measures.

Donor accountability pressures in Non-profits → strong transparency & audit orientation

Pattern observed: Non-profit respondents scored strongly on transparency and auditing; audit trails and documentation for donors were common.

Retroductive mechanism: Non-profits operate within layered accountability regimes: they must satisfy beneficiaries, regulators, and most immediately, donors and funders. Donors often require rigorous documentation and measurable outcomes. This institutional pressure creates a **compliance mechanism** that channels organizational behavior toward traceability and auditability. As a result, Non-profits adopt logging and audit practices not only to support ethical HITL use but also to meet funder requirements. This institutional logic strengthens Transparency & Auditing scores, even when contextual adaptation or technical capacity remains partial.

Evidence link: Respondents explicitly referred to audit trails for donor reporting and program-officer approval; these constraints explain why Non-profits show moderate composite sufficiency despite infrastructural limits.

Decision sensitivity and cultural misalignment in Education → high impact awareness and localized human checks

Pattern observed: Education respondents emphasized irreversible consequences for learners and described concrete human override practices.

Retroductive mechanism: educational decisions (e.g., assessment, attendance consequences, disciplinary actions) affect individual life chances directly and often irreversibly. This stakes-sensitivity creates moral and social pressure on institutions (teachers, administrators) to retain human deliberation. Moreover, education operates within rich cultural contexts where local norms, festivals, family arrangements, and linguistic variation influence how data should be interpreted. When algorithmic outputs fail to capture such nuances, practitioners deploy local knowledge to correct or contextualize outcomes. Thus, a combined mechanism of stakes sensitivity + cultural embeddedness produces robust human intervention practices in Education.

Evidence link: Quotes about a student being wrongly identified during a community festival illustrate how cultural nuance necessitates human judgment. This mechanism aligns with high Impact Severity and meaningful Human Agency.

Capacity, cost, and vendor-dependence as cross-sector constraints

Pattern observed: Across sectors, training gaps, affordability, and overreliance on third-party providers emerged as recurrent barriers.

Retroductive mechanism: Limited institutional resources (budget, skills, infrastructure) produce **capacity constraints** that limit both the ability to create context-sensitive models and the competence to interpret and act on AI outputs. Vendor dependence compounds this: when organizations outsource AI functions to third parties, they often lose visibility and control over model behavior and logging practices. This combination reduces effective transparency and human agency. Even where audit logs exist, lack of capacity to interpret them renders transparency nominal rather than practical.

Evidence link: Recurrent references to training needs, costs of subscriptions, and opaque vendor logs show this mechanism operates across sectors, explaining why some transparency measures do not translate into improved HITL sufficiency.

How the findings confirm, qualify, and extend the HITL Ethical Assessment Framework

Confirmation: The four framework components capture the key dimensions practitioners consider when evaluating HITL AI. Impact sensitivity reliably predicts stronger human oversight; transparency and auditing are central to perceived ethical sufficiency; contextual ambiguity mediates whether automation is appropriate.

Qualification: The framework assumes that stronger presence of each component straightforwardly increases ethical sufficiency. Our findings qualify that assumption: institutional incentives, resource constraints, and stakeholder accountability can decouple components from ethical practice. For example, Non-profits can have strong auditing (high Transparency) yet still experience constrained contextual adaptation because donor requirements narrow permissible responses.

Extension: Empirically, emergent themes particularly Training & Capacity, Vendor Dependence, and Donor Influence operate as cross-cutting moderators that affect all four components. We therefore propose extending the framework to include a fifth layer: Capacity & Governance Context (resources, vendor relationships, and accountability regimes). This addition helps explain why similar technical measures produce different outcomes across sectors.

Practical implications

Clarify and formalize human roles. Organizations (especially in IT) should document decision responsibilities and explicitly empower named roles to override or review AI outputs. Role clarity can be embedded in standard operating procedures and escalation protocols.

Invest in interpretability + training together. Transparency efforts must be paired with investment in staff training so that logs and explanations are actionable. Training programs should be sector-tailored (e.g., teachers vs. program officers vs. system administrators).

Design for local context. HITL workflows must include mechanisms to surface cultural edge cases and incorporate local knowledge. In education, for example, systems should have procedures for teachers to flag and annotate context that informs model retraining.

Audit vendor relationships. Where third-party providers are used, contracts should require explainability, data access, and logging standards; donors and regulators can insist on these clauses to reduce opaque vendor dependence.

Align donor incentives with local adaptability. Donors should design accountability frameworks that encourage contextual adaptation rather than rigid KPIs that incentivize algorithmic standardization.

CONCLUSION

This study examined Human-in-the-Loop (HITL) AI in Education, IT, and Non-profit sectors using the HITL Ethical Assessment Framework, which we developed to evaluate ethical sufficiency through four components: Impact Severity, Contextual Ambiguity, Human Agency, and Transparency & Auditing. By applying this framework to qualitative data, we demonstrated its analytic utility in capturing sectoral differences and diagnosing ethical strengths and weaknesses.

Our findings reveal that Education emphasizes human oversight due to high decision stakes and cultural nuance, Non-profits prioritize transparency under donor accountability, while IT organizations privilege efficiency at the expense of agency. Retroductive analysis identified underlying mechanism's institutional role ambiguity, donor pressures, cultural misalignment, and capacity constraints that shape how HITL principles are enacted.

The study's contribution is twofold: empirically validating and extending the HITL framework with a Capacity & Governance Context dimension, and practically offering actionable recommendations for policy and organizational practice.

REFERENCES

1. Berente, N., et al., Managing Artificial Intelligence. MIS Quarterly, 2021. **45**(3): p. 1433-1450.
2. Shrestha, Y.R., S.M. Ben-Menahem, and G. von Krogh, Organizational Decision-Making Structures in the Age of Artificial Intelligence. California Management Review, 2019. **61**(4): p. 66-83.
3. Wamba, S.F., Akter, S., & Guthrie, C., Making big data analytics perform: the mediating effect of big data analytics dependent organizational agility. . Systèmes d'information & management, , 2020. **25(2), 7-31.**
4. Lyytinen, K., J.V. Nickerson, and J.L. King, Metahuman systems = humans + machines that learn. Journal of Information Technology, 2020. **36**(4): p. 427-445.
5. Jarvenpaa, S.L., Sourcing data for data-driven applications: foundational questions. In Research handbook on artificial intelligence and decision making in organizations. Edward Elgar Publishing., 2024: p. (pp. 17-37). .
6. Owoc, M.L., Sawicka, A., & Weichbroth, P. , Artificial intelligence technologies in education: benefits, challenges and strategies of implementation. In IFIP international workshop on artificial intelligence for knowledge management. 2019: p. (pp. 37-58).
7. Azeem, S. and M. Abbas, Personality correlates of academic use of generative artificial intelligence and its outcomes: does fairness matter? Education and Information Technologies, 2025. **30**(13): p. 18131-18155.
8. Amershi, S., et al., Guidelines for Human-AI Interaction, in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019. p. 1-13.
9. Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). , Automation bias: a systematic review of frequency, effect mediators, and mitigators. Journal of the American Medical Informatics Association, , 2012. **19(1),** : p. 121-127.
10. Abdel-Karim, B., et al., How AI-Based Systems Can Induce Reflections: The Case of AI-Augmented Diagnostic Work. MIS Quarterly, 2023. **47**(4): p. 1395-1424.
11. Teodorescu, M.H.M., et al., Failures of Fairness in Automation Require a Deeper Understanding of Human–ML Augmentation. MIS Quarterly, 2021. **45**(3): p. 1483-1500.
12. Yin, R.K., Case study research and applications. Thousand Oaks, CA: Sage., 2018. **(Vol. 6).** .
13. Danermark, B.A., Interdisciplinary research and critical realism the example of disability research. . 5(1), , 2002. **56-64.**
14. Sawyer, B.D., Miller, D. B., Canham, M., & Karwowski, W., Human factors and ergonomics in design of a 3: automation, autonomy, and artificial intelligence. Handbook of human factors and ergonomics, , ed. 1385-1416. 2021.
15. Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. , Power to the people: The role of humans in interactive machine learning. AI magazine, , 2014. **35(4),**: p. 105-120.
16. Storey, V.C., Yue, W. T., Zhao, J. L., & Lukyanenko, R. (2025). , Generative artificial intelligence: Evolving technology, growing societal impact, and opportunities for information systems research. . Information Systems Frontiers, , 2025: p. 1-22.
17. Doshi-Velez, F., & Kim, B., Towards a rigorous science of interpretable machine learning. . arXiv preprint arXiv:, 2017. **1702.08608.**
18. Jiang, N., Liu, X., Liu, H., Lim, E. T. K., Tan, C. W., & Gu, J., Beyond AI-powered context-aware services: the role of human–AI collaboration. Industrial Management & Data Systems, , 2023. **123(11), 2771-2802.**
19. Bryson, J.J. and A. Theodorou, How Society Can Maintain Human-Centric Artificial Intelligence, in Human-Centered Digitalization and Services. 2019. p. 305-323.
20. Rahwan, I., Society-in-the-loop: programming the algorithmic social contract. . Ethics and information technology, , 2018. **20(1), 5-14.**

Ethical Considerations and Approval

This study was conducted in full compliance with established ethical standards for research involving human participants. Ethical approval was obtained from the Institutional Review Board of the American University of Nigeria; after undergoing the ethical certification review with the certificates number Record ID 3185315 and Record ID 37009720 and informed consent was secured from all participants prior to data collection.

Conflict of Interest

The authors declare that there are no conflicts of interest, financial or otherwise, that could have influenced the conduct or outcomes of this research.

Data Set Availability

The datasets generated and analyzed during the current study are available and will be made available when needed.