e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030109

Embedding Human Oversight in Semi-Automated Decision Structures for Critical Applications

Ajay Kumar Suwalka[1], Nirmal Singh[1]*, Awanit Kumar[2]

Assistant Professor, Department of Computer Science & Engineering, Sangam University, Bhilwara, Rajasthan, India[1]

Associate Professor, Department of Computer Science & Engineering, Career Point University, Kota, Rajasthan, India[2]

Abstract: With more and more computer-based systems used in sectors like medicine, banking, defense, and the judiciary, it becomes clear that it is very important to have continued human participation. This article explores the content, benefits and drawbacks of decision systems which maintain a clear role for human judgment particularly in high-stakes outcomes. Through examples based on concrete system implementations and by introducing a modularized model of the system, we demonstrate how blending human oversight with automated recommendations increases accountability, reduces operational errors, and gives voice to ethical responsibility. Focus Items: Iterative feedback, trust optimization and role clarity in the decision-making lifecycle. This work offers a framework to abstract the structured oversight of humans within core decision workflows associated with critical decisions and is meant to inform work in operational efficiency and governance.

Keywords*:* Automated decision-making, human supervision, Accountability, Ethical responsibility, Trust optimization, Iterative feedback

1. INTRODUCTION

The scope of automation has been widely expanded in a variety of industries due to technological advances. However, life conditions about health or freedom or high value assets might not be based on automated reasoners alone. Instead, hybrid systems such as the one that pair algorithmic suggestion with human judgment are becoming a safer and more accountable way forward. This study contributes to the literature by critically examining these frameworks across various industrial contexts and assess their dependability, ethical positioning and operational soundness in mission-critical environments. Addressing this, we present a four-layer model that involves data acquisition, intermediate processing, human decision oversight and a feedback loop to maintain human responsibility while preserving computational efficiency. This paper presents a conceptual framework of semi-automated decision systems, and considers its applicability across diverse high-stakes domains in a theoretical way.

2. BACKGROUND AND MOTIVATION

2.1 The Imperative for Oversight

In problem-sensitive settings such as emergency medicine, finance (compliance), or aviation control the consequences of even a small mistake can be huge. People also need to make moral judgments, contextual interpretations, and decisions.

2.2 Limitations of Fully Automated Systems

You can see why the recent string of events makes it clear how dangerous it is to fully automate sensitive areas.For example, the COMPAS system has been attacked in US court for giving different risk scores based on race (Angwin et al, 2016).The epic system emergency triage software is also said to be giving COVID-19 patients too low a priority (Heaven, 2020).These are examples of moral and operational gaps in systems that don't have people watching over them. [11]
Some important problems with systems that don't have oversight include Not being able to understand Too much dependence

- Lack of interpretability
- Excessive dependence on outputs
- Insufficient error resilience
- Unclear ethical accountability

These factors strengthen the case for architectures that maintain a role for human discretion and judgment. [7]

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030109

3. Structured Model for Oversight-Based Decision Systems

A layered model is proposed, supporting continuous interaction between system-generated outputs and human evaluators. While this framework introduces valuable ethical safeguards, it may increase decision latency, necessitating domain-specific calibration.

3.1 Data Acquisition Layer

The integrating of raw data from such input-sources, ranging from sensors to text reports, forms this phase.

3.2 Processing Layer

At that level the tool algorithms gesture to trends or decision-level detail, and show signs of potential classifications or recommendations.

3.3 Decision Oversight Layer

Human agents will then decide how to handle the recommendations, and whether they should accept, modify or reject them based on their professional knowledge, context and values (either in isolation or in a group).

3.4 Learning and Feedback Mechanism

We track and analyze human interventions to improve the future behavior of the system, which will be more aligned with what experts want.

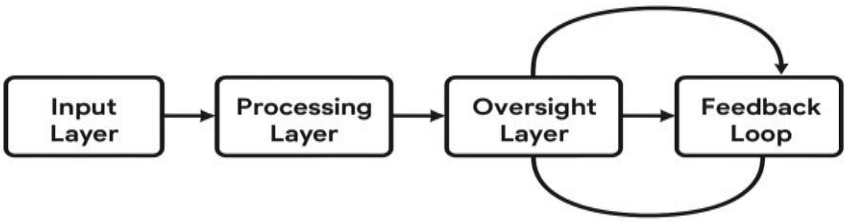

Fig.1: Layered Decision Systems

4. Practical Implementations

4.1 Medical Prioritization Systems

The emergency department, semi-automated systems suggest urgency based on symptoms. These recommendations can be overridden by clinicians given the subtlety of asymptomatology.

4.2 Surveillance in Finance

Surveillance platforms identify transactions that are tagged as suspect. These are then reviewed by human analysts, who assess a range of behavioral, legal and contextual elements to determine whether they should be escalated.

4.3 Legal Risk Scoring

Risk assessment instruments provide quantitative estimates in bail and parole decisions. But the judges - who have the final say should weigh these suggestions against a wider sweep of legal and social evidence.

One of the most controversial cases is when COMPAS was used to predict judicial parole decisions, a task in which it exhibited predictive failures and racial inequalities (Angwin et al., 2016). In both of the examples, courts should be hesitant to adopt or depart from scores as a distinct consideration, let alone one whose calculations are often opaque. [5]

TABLE NO.1: CROSS-MAPPING OF LAYERED DECISION SYSTEMS

Sr. No.	Domain	Layer	Process Description	Observed/Omitted Oversight	Consequences
1	**Medical Triage**	Data Acquisition	Patient vitals and symptoms entered into the triage interface	Incomplete or noisy symptom entry (e.g., during high workload)	Incorrect urgency ranking; delayed care; risk to patient safety
		Processing	Algorithm calculates	Black-box scoring	Over-reliance on

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030109

		Layer	priority score based on rules or data-driven criteria	without explaining rationale to physicians	system; potential neglect of critical indicators
		Decision Oversight	Physician reviews and confirms or overrides recommendations	Lack of contextual patient history or unavailability of staff	Unquestioned acceptance of machine output under time pressure
		Feedback Mechanism	Physician corrections not logged or analyzed	No system learning; repeated misclassification patterns	Diminished accuracy over time; institutional liability
2.	**Financial Surveillance**	Data Acquisition	Transaction records, metadata, customer risk profiles	Inaccurate tagging or missing customer behavioral context	False positives or missed anomalies in fraud detection
		Processing Layer	System flags transactions using rule-based or statistical models	Over-flagging due to threshold rigidity	Alert fatigue; increased manual workload; reduced trust in system
		Decision Oversight	Compliance officer evaluates flagged events	Review based on limited socio-economic or legal context	Escalation of harmless activity or missing of nuanced fraud
		Feedback Mechanism	Analyst feedback on false positives not used to update models	No adaptive refinement; recurring false flags	Erosion of analyst trust; inefficiency
3.	**Judicial Risk Scoring**	Data Acquisition	Demographic and criminal history data compiled	Incomplete social, psychological, or environmental data	Biased or misleading risk scores
		Processing Layer	Tool outputs recidivism score (e.g., COMPAS)	Lack of transparency in algorithmic logic	Bias amplification; perceived injustice; public backlash
		Decision Oversight	Judge considers system output before ruling on bail/parole	Legal over-reliance or systemic pressure to accept scores	Ethically questionable rulings; reduced judicial discretion
		Feedback Mechanism	Judges' overrides not tracked systematically	No judicial learning loop; no audit trail	Policy stagnation; repeat systemic error

Table No. 1 provides a cross-mapping of the proposed four-layer model to the case studies discussed. It highlights how oversight lapses in each layer can lead to operational, ethical, or legal consequences.

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030109

5. Building Trust in Hybrid Decision Environments

A recent idea in the behavioral-system interaction in general, and HCI specifically either trust elasticity that captures the dynamic changes of user trust in semi-automated systems over time or concept through repeated exposures to the system, degree of transparentity of the system design, or recent performance trends. In contrast to the binary trust/distrust relations, the notion of trust elasticity/discipline captures how users dynamically adjust their confidence after successful predictions and observed confusion. For example, the reliable accuracy of a system that is not life critical can establish early trust; but one catastrophic failure can suddenly erode it particularly in high-stakes settings. Systems need to be designed that dynamize over and reflect trust elasticity by providing confidence metrics, context-aware explanations, and user-centered override logs in order to avoid the twin pitfalls of over-reliance on automation and rejection specificity. Future models of trust should consider not only technical reliability, but also Enable factors which act to reinforce the psychological resilience of the user against flawed, yet assisting systems.
Studies in behavioral science confirm that automation bias may cause the users to over-rely on system outputs (Goddard et al., 2012). To do so, hybrid systems should provide confidence intervals, decision rationales and override histories to ensure human reviewers maintain their focus and skepticism. The dashboards to measure and visualize user engagement, and correction ratio could be one application in future. [2]
A critical issue in such systems is automation complacency where users place undue confidence in outputs. Maintaining balanced trust is essential to ensure critical review and prevent blind acceptance. Strategies to support this include:

- Clear representation of supporting data
- Indicators reflecting the reliability of outputs
- Structured training programs for users
- Routine performance evaluations

6. Legal and Ethical Implications

Regulations like Article 22 of the EU GDPR emphasize that individuals shouldn't be influenced by decisions based solely on automated processing. For healthcare, U.S. FDA regulations regarding SaMD (Software to be used as Medical Device) require the traceability of clinical data and a valid medical diagnosis and thereby guaranteeing the human interpretationability. These frameworks are a reflection of the emerging consensus that decision-making systems must to be accountable, transparent and equitable.
At the end human oversight is the key to ethical reassurance, however expecting end-users or individuals to be able to spot and rectify all system flaws is an unreasonable responsibility for these individuals particularly under stressful circumstances. Ethical Fattening. If there is a general awareness of the different circumstances is present ethics, then the robustness of ethical conduct is more institutional than individual and can be public. This is why it is necessary to implement organizational accountability measures such as formal oversight, ethical audit procedures reviewing for bias effects and the statutory records of override decision-making. Thirdly, ethical review boards and oversight mechanisms are able to take action whenever they are in possession of evidence that shows that the system in general is putting quality over quantity when it comes to the performance of its direct duties. In this age of huge data, the ability to prevent embarrassment in the ethical sphere could allow an institution to work more transparently and increase public's trust in sensitive areas such as health care delivery, social welfare, or criminal justice in cases where decisions are based on fair treatment of the citizens and the social justice system.

For shared ownership schemes T&Cs need to be bullet-proof. Key considerations include:

- **Accountability Clarity:** Accountability Knowing who (system designer, human reviewer or both) will be held accountable.
- **Bias Control**: Keeping Bias under Control Taking out patterns that are biased, which are commonly caused by bad data or feedback tours.
- **User Transparency:** Clear information for users People Being open about how decisions are made, especially when it comes to human rights, is what transparency means.

7. Forward-Looking Recommendations

To ensure that oversight-based decision-making processes work well and are honest, future research and implementation efforts should focus on both short-term operational improvements and long-term strategic goals.

7.1 Short-Term Priorities

- Find out how user-generated content changes what people do and decide, such as the unbiased rates of overriding behavior, the accuracy of corrections, and the time it takes to analyze.

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030109

- Make sure your escalation policies are aware of the domain, and always be on the lookout for situations that pose a risk and need human action.
- Make interfaces that can change and react to input and user experience loops. People who review can keep in the loop this way instead of merely checking off boxes.

7.2 Long-Term Priorities

- Use governance models that include people from other fields, like legal, ethical, technological, and domain expertise.
- Create systems that can be tested to ensure that both machines and people can understand and follow the logic and likely paths.
- Ask for a curriculum and work on establishing and putting into place moral rules, especially for businesses and how the government buys things.

7.3 Standardization Efforts

It could be a good idea to set a worldwide standard for demand and benefit (and beyond) in order to get the most out of hybrid decision systems, as was said in the introduction. ISO/IEC JTC1 SC42 AI and other similar initiatives may have started to set rules for AI.. regarding trustworthiness and risk reduction. Through this adopting and implementing new standards, such as these, can help companies evaluate how they are performing to the best practices in the UK as well as internationally as well as guide the development of services that anticipate the future regulations. Trust and the foundations for an international platform for the ethics of steering systems are built by being part of such ecosystems.
It is necessary to further investigate the issue, to create quantifiable tools with that we can evaluate the effectiveness and efficiency of human stewardship within this system. This includes calculating rate of overrides and accuracy in corrections and the time it takes to make a decision. Additionally, escalation rules specific to a particular service could also be used in conjunction, and an conclusion to trigger the human-in-the-loop process, even if not doing it could result in a the worst outcome (e.g. disqualifying medical alerts or legal judgments against minors). Additionally, there would need to be the creation of a flexible interface that learns from previous experiences and users according to the maturity and knowledge degree of the reviewers.

Future Research should emphasize:

- A quantitative assessment of the human contribution to the system
- Improve the type of task and translations on the Help pages.
- Intelligent user interface that is able to learn from experience and knowledge acquired through usage habits

8. CONCLUSION

Indeed, the more we rely on automatic systems, the greater our need for principled decision making within context. But it's not just protection against idiocy we should be baking into this. It is necessary for fairness, interpretability and robustness. With clean transparency and accountability lines, open discussions regarding where in the process chain will setting stages transparent help come and relax review processes which "doesn't require an institution of power to press +1 on a rule" we should be looking at making systems that define us by our human worth that can leverage technology to work better to allow sensemaking how 2 get'uh done.
And when the stakes for outcomes are high, autonomous systems won't be very useful if they cannot make complex and explainable judgments. In this paper we are proposing an integrated approach in which computational methods are complementary rather than competitive to human judgment. But built and stewarded responsibly, such systems could offer us a way to respond to uncertainty with efficacy and ethical certainties.
And yet with decision systems making inroads into high-stakes industries, to exclude the exercise of human judgment from these loops is a failure in operation — and has an ironic corollary: It undermines confidence in the most critical infrastructure that we have. (opens in new tab)Many a not small-entity has placed its authority on evidence, courts and hospitals and NSC-8s as financial systems. Eple that. 1 ∘ d 13 Some transparency and fair handling and hum an empathy not always given by the full-automatised systems. Inject that human touch and supervision back into the automated systems, and all of a sudden the situation can be a lot more subtle: now you have awareness-of-situation front-and-center as a driver of decision. The only way to do that it for them, tandem with the businesses themselves, to formalize their oversight, roles and feedback loops using language-invariant processes that lead us towards systems not just efficient but socially and ethically possible.

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030109

REFERENCES

[1] Parasuraman, R., & Riley, V., "Human and Automation: Use, Misuse, Disuse, Abuse", Human Factors, vol. 39, no. 2, pp. 230–253, 1997.
[2] Goddard, K., Roudsari, A., & Wyatt, J. C., "Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators", Journal of the American Medical Informatics Association, vol. 19, no. 1, pp. 121–127, 2012.
[3] Amershi, S., et al., "Power to the People: The Role of Humans in Interactive Systems", Communications of the ACM, vol. 57, no. 11, pp. 86–94, 2014.
[4] European Union, "General Data Protection Regulation (GDPR), Article 22", Official Journal of the European Union, vol. L119, pp. 1–88, 2016.
[5] Angwin, J., Larson, J., Mattu, S., & Kirchner, L., "Machine Bias", ProPublica, May 23, 2016.
[6] Doshi-Velez, F., & Kim, B., "Towards a Rigorous Science of Interpretable Machine Learning", arXiv preprint arXiv:1702.08608, 2017.
[7] Lipton, Z. C., "The Mythos of Model Interpretability", Communications of the ACM, vol. 61, no. 10, pp. 36–43, 2018.
[8] Selbst, A. D., & Barocas, S., "The Intuitive Appeal of Explainable Machines", Fordham Law Review, vol. 87, no. 3, pp. 1085–1139, 2018.
[9] Raji, I. D., & Buolamwini, J., "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Systems", Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, pp. 429–435, 2019.
[10] Miller, T., "Explanation in Artificial Intelligence: Insights from the Social Sciences", Artificial Intelligence, vol. 267, pp. 1–38, 2019.
[11] Heaven, W. D., "Hundreds of AI Tools Have Been Built to Catch COVID. None of Them Helped", MIT Technology Review, June 2020.
[12] U.S. Food and Drug Administration (FDA), "Software as a Medical Device (SaMD): Clinical Evaluation Guidance for Industry", 2021.