European Journal of Computer Science and Information Technology, 13(51),27-31, 2025
Print ISSN: 2054-0957 (Print)
Online ISSN: 2054-0965 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

Humans in the Loop, Lives on the Line: AI in High-Risk Decision Making

Merlin Balamurugan
Vice President, Leading Banking Organization

Kumar Shanmugasamy
Senior Vice President, Leading Banking Organization

Saranya Balaguru
Solution Architect, Leading Healthcare Organization

doi: https://doi.org/10.37745/ejcsit.2013/vol13n512731 Published July 14, 2025

Citation: Balamurugan M., Shanmugasamy K. and Balaguru S. (2025) Humans in the Loop, lives on the Line: AI in High-Risk Decision Making, *European Journal of Computer Science and Information Technology*, 13(51),27-31

Abstract: *In high-stakes arenas like healthcare, finance, and fraud detection, the cost of AI getting it wrong can be immense, from ethical fallout to real-world harm. This paper explores how Human-in-the-Loop (HITL) artificial intelligence offers a crucial safeguard in such scenarios by keeping humans actively involved in decision-making processes. We dive into how human oversight can make AI systems more accountable, interpretable, and adaptable to complex, real-world challenges. Through real-world case studies and established frameworks, we assess how HITL can curb algorithmic bias, promote fairness, and deliver better outcomes. But it's not all smooth sailing: issues like cognitive overload, ambiguous roles, and scaling challenges are also part of the equation. To move forward, we outline key design principles and propose concrete evaluation metrics aimed at building HITL systems we can truly trust.*

Keywords: human-in-the-loop (HITL), explainable AI (XAI), high-risk decision-making, AI ethics and accountability, human-AI collaboration

INTRODUCTION

Artificial intelligence is no longer confined to research labs or recommendation engines — it's now helping diagnose illnesses, flag financial fraud, and influence decisions in courtrooms and banks. But when lives, livelihoods, or justice are on the line, handing over full control to machines raises serious ethical and practical concerns. Despite their speed and scale, fully autonomous AI systems often fall short in areas where human judgment, context, and accountability matter most.

Enter the Human-in-the-Loop (HITL) approach — a hybrid model where humans remain actively involved in the AI decision-making process [1]. By blending algorithmic power with human insight, HITL systems aim to offer the best of both worlds: the precision of machines and the wisdom of people. This paper explores how HITL can be designed and deployed effectively in high-risk settings such as medical diagnosis, fraud detection, and loan approvals. We examine not just the benefits — increased fairness [2], trust, and

adaptability — but also the challenges of striking the right balance between automation and human oversight. Finally, we propose actionable frameworks to guide the development of HITL systems that are not only high-performing but also ethically sound and socially responsible [3].

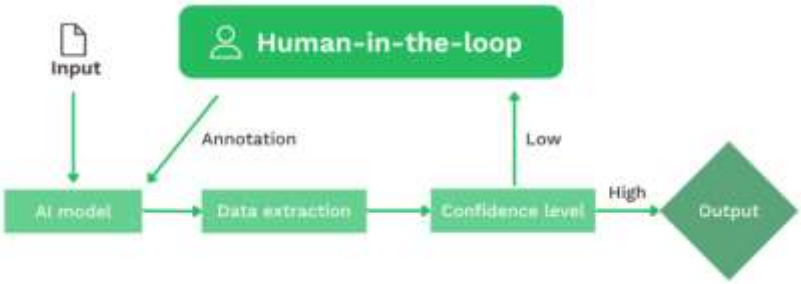

Figure 1: Human-in-the-Loop Annotation [14]

I.Problem statement

The following problems highlight key areas of concern [4]:

- **Lack of Trust and Transparency in Autonomous AI Systems:** In high-risk domains such as healthcare and finance, fully automated AI systems often operate as "black boxes," making it difficult for stakeholders to interpret or trust their decisions.
- **Insufficient Handling of Rare and Complex Scenarios:** AI models frequently underperform in edge cases that require nuanced human judgment, increasing the risk of harmful or suboptimal outcomes in critical situations.
- **Ambiguity in Human-AI Interaction Roles:** There is a lack of clear guidelines regarding the division of responsibilities between human operators and AI systems, leading to inefficiencies, delayed interventions, and accountability challenges.
- **Scalability Constraints on Human Oversight:** Maintaining consistent human oversight across large-scale or real-time AI deployments presents significant operational challenges and may undermine the advantages of automation if not effectively managed.
- **Unmitigated Bias and Ethical Risks [5]:** In the absence of human review, AI systems are more likely to perpetuate or exacerbate bias, resulting in unfair or ethically questionable decisions, particularly in sensitive areas such as credit scoring and fraud detection.

SOLUTION

To address the identified challenges, the following solutions are proposed to enhance the effectiveness, accountability, and ethical integrity of Human-in-the-Loop AI systems [6]:

- **Define Clear Human-AI Interaction Protocols:** Establish structured guidelines outlining the roles, responsibilities, and decision thresholds for human involvement. This ensures effective collaboration, enhances accountability, and reduces ambiguity in AI-assisted decision-making processes.
- **Adopt Tiered Risk-Based Decision Frameworks:** Implement hierarchical decision systems in

which routine, low-risk decisions are handled autonomously, while high-risk or uncertain cases are escalated for human evaluation. This approach optimizes efficiency while safeguarding critical outcomes.

- **Integrate Explainable AI (XAI) Mechanisms [7]:** Develop and deploy AI models capable of generating transparent, interpretable justifications for their outputs. These explanations support human reviewers in understanding, validating, or contesting AI-driven decisions.
- **Enhance Human Oversight Through Training and Interface Design [8]:** Provide human operators with comprehensive training, contextual information, and intuitive user interfaces to empower informed and confident decision-making in conjunction with AI systems.

- **Establish Continuous Human Feedback Loops:** Incorporate mechanisms for collecting and analyzing human feedback to refine AI models over time. This helps detect model drift, mitigate bias, and adapt to evolving scenarios or edge cases.
- **Optimize Oversight Through Selective Automation [9]:** Utilize intelligent automation to prioritize and flag only cases requiring human attention. This enables scalable oversight by focusing human resources where they are most needed, without compromising safety or quality.
- **Define Ethical and Performance Standards:** Develop and apply standardized benchmarks aligned with regulatory and ethical guidelines to evaluate HITL AI systems. Metrics should include fairness, transparency, accuracy, and user satisfaction across real-world use cases.

II.Benefits of solutions

Implementing Human-in-the-Loop (HITL) AI in high-risk decision-making presents substantial advantages across critical domains, including healthcare [10][12], finance, and justice. These systems leverage the complementary strengths of human insight and machine processing [11]. By incorporating human judgment, HITL AI enhances the accuracy of decisions and helps prevent potentially harmful errors. The presence of human oversight also contributes to greater transparency in AI operations. This, in turn, fosters trust among stakeholders, particularly in contexts that involve ethical considerations. Moreover, human feedback enables AI systems to adapt continuously to changing conditions and complex real-world situations. Such adaptability ensures the relevance and reliability of AI outcomes over time. Ultimately, HITL frameworks offer a balanced model that supports fairness, accountability, and safety in the deployment of AI technologies.

III.Conclusion

Human-in-the-Loop (HITL) AI presents a compelling approach for enhancing the safety, fairness, and reliability of decision-making in high-risk domains [13]. Key takeaways include:

Balanced Decision-Making: HITL systems effectively combine human reasoning with AI's speed and scalability, leading to more reliable and context-aware outcomes.

Improved Trust and Transparency: Human oversight enhances stakeholder confidence by enabling decisions that are explainable, auditable, and open to challenge.

Mitigation of Bias and Ethical Risks: Active human monitoring helps identify and correct biased or ethically questionable outputs, particularly in sensitive sectors like healthcare and finance.

Scalable Oversight Through Smart Design: Intelligent frameworks, such as triage-based decision models, enable efficient human involvement without sacrificing the benefits of automation.

Need for Continued Research and Standards: Advancing HITL systems requires ongoing research, the development of standardized evaluation metrics, and alignment with ethical and regulatory standards.

REFERENCES

[1] R. Munro, Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI, 1st ed. Shelter Island, NY, USA: Manning Publications, 2021.
[2] B. Christian, The Alignment Problem: Machine Learning and Human Values, 1st ed. New York, NY, USA: W. W. Norton & Company, 2020.
[3] B. Shneiderman, Human-Centered AI, 1st ed. Oxford, UK: Oxford University Press, 2022.
[4] A. Kothari, Designing Human-Centric AI Experiences, 1st ed. Sebastopol, CA, USA: O'Reilly Media, 2020.
[5] J. Smith, Ethics and AI: Principles and Practice, 1st ed. London, UK: Routledge, 2023.
[6] L. Zhao and M. Taylor, Trustworthy AI Systems: Design, Evaluation, and Regulation, 1st ed. New York, NY, USA: Springer, 2022.
[7] W. Samek, A. Taly, and K.-R. Müller, Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, 1st ed. Cham, Switzerland: Springer, 2019.
[8] C. O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, 1st ed. New York, NY, USA: Crown Publishing Group, 2016.
[9] R. Benjamin, Race After Technology: Abolitionist Tools for the New Jim Code, 1st ed. Cambridge, UK: Polity Press, 2019.
[10] A. Bohr and K. Memarzadeh, Eds., Artificial Intelligence in Healthcare, 1st ed. Cambridge, MA, USA: Academic Press, 2020.
[11] S. Gupta and H. Lee, Human-AI Collaboration: Models and Methods, 1st ed. Berlin, Germany: De Gruyter, 2021.
[12] D. A. Clifton, Ed., Machine Learning for Healthcare Technologies, 1st ed. Stevenage, UK: The Institution of Engineering and Technology (IET), 2016.
[13] M. Rivera, Bias in AI Systems: Detection and Mitigation Strategies, 1st ed. San Francisco, CA, USA: Morgan Kaufmann, 2023.
[14] https://www.klippa.com/en/blog/information/human-in-the-loop/

Author Profile

Merlin Balamurugan is a distinguished Cognitive Engineer with 18 years of specialized experience in Digital Identity, Banking, and Finance. She has adeptly managed numerous projects, integrating Artificial Intelligence and diverse Banking methodologies. In her role, Merlin has provided strategic leadership in navigating complex issues and ensuring alignment with organizational objectives. She has also played a pivotal role in contributing thought leadership to the strategic planning process.

Kumar Shanmugasamy is a seasoned Payment Lead with over 18+ years of experience in the financial industry. He has worked with top-tier companies, leading initiatives to optimize payment processes, enhance security, and streamline transaction workflows. Kumar holds a bachelor's degree in computer science and engineering from Madras University, Chennai, India. His expertise includes implementing cutting-edge payment solutions, driving digital transformation using AI, and managing cross-functional teams.

Saranya Balaguru is a cognitive automation professional with over 10 years of experience specializing in the healthcare industry. She has led numerous projects integrating technologies like Robotic Process Automation (RPA) and Generative AI to streamline operations and enhance productivity across various business units. Saranya holds a master’s degree in computer science and engineering from Anna University, Chennai, India, and has contributed to industry publications in collaboration with the head of her department at the university. Her expertise lies in leveraging advanced automation to drive efficiency and innovation in healthcare.