

Novelty Assessment Report

Paper: Formalising Human-in-the-Loop: Computational Reductions, Failure Modes, and Legal-Moral Responsibility

PDF URL: <https://openreview.net/pdf?id=KR8viVTrX4>

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-29

Abstract

We use the notion of oracle machines and reductions from computability theory to formalise different Human-in-the-loop (HITL) setups for AI systems, distinguishing between trivial human monitoring (i.e., total functions), single endpoint human action (i.e., many-one reductions), and highly involved human-AI interaction (i.e., Turing reductions). We then proceed to show that the legal status and safety of different setups vary greatly. We present a taxonomy to categorise HITL failure modes, highlighting the practical limitations of HITL setups. We then identify omissions in UK and EU legal frameworks, which focus on HITL setups that may not always achieve the desired ethical, legal, and sociotechnical outcomes. We suggest areas where the law should recognise the effectiveness of different HITL setups and assign responsibility in these contexts, avoiding human 'scapegoating'. Our work shows an unavoidable trade-off between attribution of legal responsibility, and technical explainability. Overall, we show how HITL setups involve many technical design decisions, and can be prone to failures out of the humans' control. Our formalisation and taxonomy opens up a new analytic perspective on the challenges in creating HITL setups, helping inform AI developers and lawmakers on designing HITL setups to better achieve their desired outcomes.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Formalising Human-in-the-Loop Setups Using Computational Reductions**

A total of **25 papers** were analyzed and organized into a taxonomy with **20 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Theoretical Foundations and Formal Frameworks**
- **Interactive System Design and Optimization**
- **Machine Learning with Human Feedback**
- **Dimensionality Reduction and Reliability**
- **Application Domains**
- **Supporting Methodologies**

Complete Taxonomy Tree

- Formalising Human-in-the-Loop Setups Using Computational Reductions Survey Taxonomy
- Theoretical Foundations and Formal Frameworks
 - Computational Reduction Models for Human-AI Interaction ★ (2 papers)
 - [0] Formalising Human-in-the-Loop: Computational Reductions, Failure Modes, and Legal-Moral Responsibility (Anon et al., 2026) [View paper](#)
 - [11] Can Humans Be out of the Loop? (J Zhang, 2022) [View paper](#)
 - Interaction Protocols and Decision Frameworks (2 papers)
 - [5] Relevance-Driven Decision Making for Safer and More Efficient Human Robot Collaboration (Xiaotong Zhang, 2024) [View paper](#)
 - [7] Tractable Agreement Protocols (Natalie Collina, 2024) [View paper](#)
 - Safety and Reliability Frameworks (2 papers)
 - [12] Using Formal Frameworks to Reduce Mode Confusion in Cyber-Physical Human-in-the-Loop Systems (Byrnes, 2023) [View paper](#)
 - [14] Fault Detection and Human Intervention in Vehicle Platooning: A Multi-Model Framework (Pirani, 2025) [View paper](#)
- Interactive System Design and Optimization
 - Interactive Exploration and Visualization Tools (3 papers)
 - [6] Smart computational exploration of stochastic gene regulatory network models using human-in-the-loop semi-supervised learning (Fredrik Wrede, 2018) [View paper](#)
 - [9] DendroTweaks: An interactive approach for unraveling dendritic dynamics (Roman Makarov, 2024) [View paper](#)
 - [16] Towards Interactive Pattern Search in Massive Graphs (Tahsin Reza, 2020) [View paper](#)
 - Point-of-Care and Real-Time Optimization (1 papers)
 - [10] Human-in-the-loop optimization of transcranial electrical stimulation at the point of care: a computational perspective (Yashika arora, 2022) [View paper](#)
 - Model Development Workflows and Frameworks (1 papers)
 - [21] GAMES: A dynamic model development workflow for rigorous characterization of synthetic genetic systems (Kate E. Dray, 2021) [View paper](#)
- Machine Learning with Human Feedback
 - Reinforcement Learning from Human Feedback (1 papers)
 - [18] RRHF: Rank Responses to Align Language Models with Human Feedback without tears (YUAN Zheng, 2023) [View paper](#)
 - Off-Policy Learning and Evaluation (1 papers)

- [17] Off-Policy Evaluation and Learning from Logged Bandit Feedback: Error Reduction via Surrogate Policy (Xie Yuan, 2018) [View paper](#)
- Interactive Code Generation and Verification (2 papers)
- [3] LLM-based Interactive Code Generation: Empirical Evaluation (Danil Shaikhelislamov, 2024) [View paper](#)
- [23] Interactive Code Generation via Test-Driven User-Intent Formalization (Lahiri, 2022) [View paper](#)
- Aesthetic and Preference Learning (1 papers)
- [15] Deep learning of individual aesthetics (Jon McCormack, 2020) [View paper](#)
- Dimensionality Reduction and Reliability
 - Reliability Analysis in Visual Analytics (1 papers)
 - [2] Unveiling high-dimensional backstage: A survey for reliable visual analytics with dimensionality reduction (Hyeon Jeon, 2025) [View paper](#)
 - Fast Dimensionality Reduction Algorithms (1 papers)
 - [22] Linear basis-function t-SNE for fast nonlinear dimensionality reduction (A. Gisbrecht, 2012) [View paper](#)
- Application Domains
 - Energy Systems and Resilience (1 papers)
 - [4] Modern energy resilience studies with artificial intelligence for energy transitions (Yuekuan Zhou, 2025) [View paper](#)
 - Information Retrieval and Relevance Feedback (1 papers)
 - [8] Reduct-based result set fusion for relevance feedback in CBIR (Samar Zutshi, 2024) [View paper](#)
 - Creative and Embodied Applications (1 papers)
 - [19] PoeSpin: A Human-AI Dance to Poetry System for Movement-Based Verse Generation (Li Yihua, 2025) [View paper](#)
 - Digital Hermeneutics and Textual Interpretation (1 papers)
 - [25] AI as Interpretive Aid in Qur'anic Stylistics: Ethical Foundations for Digital Hermeneutics (H Ahmad, n.d.) [View paper](#)
- Supporting Methodologies
 - Unknown-Unknowns and Model Reduction (1 papers)
 - [1] Unknown-Unknowns in Human-Centric AI Systems (Maity, 2025) [View paper](#)
 - Feedback Formalization and Terminology (1 papers)
 - [20] Formal Notations and Terminology for Users' Feedback and Its Specialization for Interactive Fault Localization (Gergo Balogh, 2022) [View paper](#)
 - Temporal Compression for Complexity Reduction (1 papers)
 - [13] Adaptive temporal compression for reduction of computational complexity in human behavior recognition. (Haixin Huang, 2024) [View paper](#)
 - Communication Reduction via Feedback (1 papers)
 - [24] The AWGN BC with MAC Feedback: A Reduction to Noiseless Feedback via Interaction (Assaf Ben-Yishai, 2015) [View paper](#)

Narrative

Core task: formalising human-in-the-loop setups using computational reductions. The field spans a diverse set of concerns, from foundational theory to practical deployment. At the highest level, the taxonomy organizes work into six main branches: Theoretical Foundations and Formal Frameworks, which develop rigorous models and reduction-based abstractions for human-AI interaction; Interactive System Design and Optimization, which addresses interface design, user modeling, and adaptive workflows; Machine Learning with Human Feedback, covering reinforcement learning from human preferences and related training paradigms; Dimensionality Reduction and Reliability, which tackles visualization, interpretability, and robustness; Application Domains, spanning robotics, healthcare, energy systems, and beyond; and Supporting Methodologies, which provide cross-cutting techniques such as protocol design and data fusion. Representative works illustrate these themes: for instance, Humans Out Loop[11] and Formal Frameworks Mode Confusion[12] anchor the theoretical side, while RRHF[18] and LLM Interactive Code Generation[3] exemplify machine learning with feedback, and Visual Analytics Dimensionality Reduction[2] highlights interpretability challenges.

Several active lines of work reveal key trade-offs and open questions. One tension lies between formal guarantees—pursued by reduction-based frameworks that treat human input as an oracle or computational resource—and the messy realities of adaptive interfaces and noisy feedback in deployed systems. Another contrast emerges between domain-agnostic methodologies, such as dimensionality reduction techniques like Linear t-SNE[22], and domain-specific applications like Brain Stimulation Optimization[10] or Vehicle Platooning Intervention[14], each of which must reconcile general principles with specialized constraints. Within this landscape, Formalising Human-in-the-Loop[0] sits squarely in the Theoretical Foundations branch, specifically under Computational Reduction Models for Human-AI Interaction. Its emphasis on rigorous reduction-based abstractions aligns closely with Humans Out Loop[11], which also explores formal characterizations of human involvement, though the two may differ in how they model the boundary between automated and human-driven decision-making. By anchoring human-in-the-loop setups in computational complexity and reduction theory, this work provides a unifying lens that complements more empirical or application-focused studies elsewhere in the taxonomy.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Can Humans Be out of the Loop?

Authors: J Zhang, E Bareinboim | **Year/Venue:** 2022 | **URL:** [View paper](#)

Abstract

â€” with incorporating the â€”human in the loopâ€” focuses on the role â€”! Second, how could we formalize the interactive agent in this â€” the polynomial optimization reduction worked as expected, â€”!

Relationship Analysis

Both papers belong to the same taxonomy category by using computational reduction models (oracle machines, Turing reductions) to formalize human-AI interaction systems. They overlap in applying computability theory to model human involvement in automated decision-making, with both examining how human input affects system outcomes through formal computational frameworks. However, the original paper focuses on legal-moral responsibility, failure modes, and safety across different HITL setup types (trivial monitoring, endpoint action, involved interaction), while the candidate paper specifically addresses reinforcement learning scenarios where humans and agents have different perceptual capabilities, introducing counterfactual agents and budget-constrained optimization problems.

Contributions Analysis

Overall novelty summary. The paper formalizes human-in-the-loop (HITL) setups using oracle machines and computational reductions from computability theory, distinguishing trivial monitoring, single-endpoint action, and highly interactive collaboration. It resides in the 'Computational Reduction Models for Human-AI Interaction' leaf, which contains only two papers total. This leaf sits within the broader

'Theoretical Foundations and Formal Frameworks' branch, indicating a relatively sparse research direction focused on rigorous formal characterizations rather than empirical or application-driven work.

The taxonomy reveals neighboring leaves addressing 'Interaction Protocols and Decision Frameworks' (tractable protocols and agreement mechanisms) and 'Safety and Reliability Frameworks' (mode confusion and formal verification). These adjacent areas share the theoretical branch but diverge in focus: the sibling leaves emphasize decision-theoretic models and fault detection, whereas the paper's leaf concentrates on reduction-based abstractions. The taxonomy's scope notes clarify that applied implementations belong elsewhere, reinforcing that this work occupies a foundational niche distinct from domain-specific applications scattered across the 'Application Domains' branch.

Among 29 candidates examined, the three contributions—formalizing HITL via reductions (9 candidates), taxonomizing failure modes (10 candidates), and analyzing legal frameworks (10 candidates)—show no clear refutations. The limited search scope means these statistics reflect top-K semantic matches and citation expansion, not exhaustive coverage. The formalization contribution appears particularly novel given the sparse leaf population, while the failure taxonomy and legal analysis may overlap with broader human-AI interaction literature not captured in this focused search. The absence of refutable pairs suggests either genuine novelty or gaps in the candidate pool.

Based on the limited search of 29 candidates, the work appears to occupy a sparsely populated formal niche, with its reduction-based approach distinguishing it from neighboring protocol-oriented or safety-focused frameworks. The analysis cannot confirm whether larger-scale searches or domain-specific legal literature would reveal closer prior work, particularly for the legal responsibility and failure mode contributions.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Formalisation of HITL setups using computational reductions

Description: The authors introduce a novel computational framework that characterises HITL setups through oracle machines and reduction types from computability theory. This formalisation distinguishes three setup types: trivial monitoring (total functions), endpoint action (many-one reductions), and involved interaction (Turing reductions), unifying disparate HITL concepts under a consistent theoretical lens.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Oracle or Teacher? A Systematic Overview of Research on Interactive Labeling for Machine Learning

URL: [View paper](#)

Brief Assessment

Interactive Labeling Overview[35] focuses on interactive labeling for machine learning training data, examining oracle vs. teacher paradigms for human annotators. This is fundamentally different from the original paper's computational framework for characterizing HITL setups through oracle machines and reduction types from computability theory.

2. Improved Inference of Human Intent by Combining Plan Recognition and Language Feedback

URL: [View paper](#)

Brief Assessment

Intent Recognition Language Feedback[33] focuses on plan and goal recognition for assistive robots using HTNs and POMDPs for dialogue-based task tracking. It does not address the theoretical formalisation of HITL setups through oracle machines and reduction types from computability theory.

3. Human-in-the-loop active learning for goal-oriented molecule generation

URL: [View paper](#)

Brief Assessment

Goal-Oriented Molecule Generation[26] focuses on active learning for molecule generation with human feedback on property predictions, not on formalizing HITL setups through oracle machines and computational reduction types from computability theory.

4. Leveraging Oracle Digital Assistant (ODA) to Automate ERP Transactions and Improve User Productivity

URL: [View paper](#)

Brief Assessment

ODA ERP Automation[32] focuses on implementing a conversational AI chatbot for Oracle ERP transaction automation. It does not address theoretical formalisation of human-in-the-loop systems using oracle machines or computational reduction theory.

5. Addressing the data bottleneck in medical deep learning models using a human-in-the-loop machine learning approach

URL: [View paper](#)

Brief Assessment

Medical Data Bottleneck[29] focuses on practical HITL implementation in medical deep learning (active learning and interactive ML for data augmentation), not on theoretical formalisation using oracle machines and computational reductions from computability theory.

6. VOICE: Visual Oracle for Interaction, Conversation, and Explanation

URL: [View paper](#)

Brief Assessment

VOICE[28] focuses on conversational visualization for science communication using LLMs and interactive 3D models. It does not address human-in-the-loop formalization, oracle machines, or computational reduction theory.

7. Modeling Variation in Human Feedback with User Inputs: An Exploratory Methodology

URL: [View paper](#)

Brief Assessment

Modeling User Feedback Variation[30] focuses on modeling variation in human feedback for interactive reinforcement learning using simulated oracles, not on formalizing HITL setups through computational reductions or oracle machines from computability theory.

8. Towards understanding and simplifying human-in-the-loop machine learning

URL: [View paper](#)

Brief Assessment

Understanding Human-in-the-Loop ML[34] focuses on iterative ML workflow optimization and system design for human-in-the-loop development cycles, not on formalizing HITL setups through oracle machines and computational reduction theory from computability.

9. You are the only possible oracle: Effective test selection for end users of interactive machine learning systems

URL: [View paper](#)

Brief Assessment

Oracle Test Selection[31] focuses on test selection for end-user testing of machine learning classifiers, not on formalizing HITL setups using oracle machines and computational reductions from computability theory.

Contribution 2: Taxonomy of HITL failure modes

Description: The authors develop a taxonomy organised into five main failure categories (machine components, process and workflow, human-machine interface, human component, and exogenous circumstances) that systematically captures how HITL setups can fail in practice. This taxonomy connects failure modes to the different computational reduction types identified in their formalisation.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Collaborative automation in factories of the future: review and survey

URL: [View paper](#)

Brief Assessment

Collaborative Factory Automation[51] focuses on factory automation systems with spatial-visual programming and augmented reality. It does not present a taxonomy of human-in-the-loop failure modes for AI systems.

2. Context-Awareness and Interpretability of Rare Occurrences for Discovery and Formalization of Critical Failure Modes

URL: [View paper](#)

Brief Assessment

Rare Occurrences Failure Modes[52] focuses on failure detection in vision systems for automated driving, not on developing a general taxonomy of human-in-the-loop failure modes across different computational setups.

3. Design and Application of a C++ Compiler Error Solution Query Platform Integrating Large Language Models and Human-in-the-Loop Support

URL: [View paper](#)

Brief Assessment

Compiler Error Query Platform[55] focuses on C++ compiler error solutions using LLMs and error classification for educational purposes, not on taxonomies of Human-in-the-loop failure modes in AI systems.

4. Scientific Knowledge Graph Construction Needs an AI-Mediated, Scientist-in-the-Loop Workflow (A Blue Sky Paper)

URL: [View paper](#)

Brief Assessment

Scientist-in-the-Loop Knowledge Graph[50] focuses on knowledge graph construction workflows in scientific domains, not on developing taxonomies of HITL failure modes in AI systems. The candidate addresses semantic guidance and expert consultation mechanisms rather than categorizing failure modes.

5. Human information interaction, artificial intelligence, and errors

URL: [View paper](#)

Brief Assessment

Human Information Interaction Errors[49] focuses on error classification in human-information interaction contexts with AI proxies, not on developing taxonomies of Human-in-the-loop failure modes in AI systems. The paper discusses error types (mistakes, slips, active/latent errors) but does not present a structured taxonomy of HITL failure categories.

6. Reliability Assurance for AI Systems

URL: [View paper](#)

Brief Assessment

AI Reliability Assurance[47] focuses on reliability engineering for AI systems in military applications, emphasizing data pipelines, adversarial attacks, and system-theoretic process analysis. It does not present a detailed taxonomy of human-in-the-loop failure modes organized by computational reduction types as the original paper does.

7. Architecting Human-AI Systems for Effective Collaboration and Oversight: Making Sense of Human/AI in/on/Over/Under/Along the Loop

URL: [View paper](#)

Brief Assessment

Architecting Human-AI Systems[54] presents a taxonomy of human-AI control configurations (human-in/on/over/under/along-the-loop) focused on control relationships, not failure modes. The candidate does not address systematic categorization of how HITL setups fail in practice.

8. Human-in-the-loop Techniques in Machine Learning.

URL: [View paper](#)

Brief Assessment

Human-in-the-Loop Techniques[48] focuses on technical ML pipeline components (data extraction, integration, cleaning, labeling) and optimization techniques (quality improvement, cost reduction, active learning), not on systematic failure mode taxonomies for HITL systems.

9. Understanding choice independence and error types in human-ai collaboration

URL: [View paper](#)

Brief Assessment

Choice Independence Human-AI[46] focuses on human delegation decisions and error evaluation in AI collaboration tasks, not on developing taxonomies of failure modes in HITL systems. The paper examines choice independence violations and error type differentiation in prediction tasks, which is a different research focus from systematically categorizing HITL failure modes.

10. Smart Biofloc Systems: Leveraging Artificial Intelligence (AI) and Internet of Things (IoT) for Sustainable Aquaculture Practices

URL: [View paper](#)

Brief Assessment

Smart Biofloc Systems[53] focuses on AI/IoT integration in aquaculture biofloc technology, not on human-in-the-loop failure taxonomies for AI systems.

Contribution 3: Analysis of legal frameworks and responsibility trade-offs

Description: The authors analyse UK and EU legal frameworks (GDPR and EU AI Act) to identify gaps in how they address HITL requirements, and reveal an inherent trade-off: HITL setups with greater explainability (involved interactions) create responsibility gaps, while setups with clearer responsibility attribution (endpoint actions) are less transparent. They provide suggestions for improving legal frameworks to prevent humans from becoming scapegoats.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Human-in-the-Loop Robotics: Enhancing Safety and Adaptability through Interactive AI Systems

URL: [View paper](#)

Brief Assessment

Human-in-the-Loop Robotics[40] focuses on technical design, cognitive ergonomics, and operational implementation of HITL systems in robotics applications. It does not analyze legal frameworks (GDPR, EU AI Act) or examine responsibility-explainability trade-offs in HITL systems as the original paper does.

2. Requirements of high-risk AI systems: AI Act. Article 14. Human oversight

URL: [View paper](#)

Brief Assessment

AI Act Human Oversight[39] focuses on technical requirements for human oversight in high-risk AI systems under EU regulation, not on analyzing responsibility trade-offs or gaps in legal frameworks regarding HITL explainability versus accountability.

3. Autonomous AI, smart seaports, and supply chain management: Challenges and Risks

URL: [View paper](#)

Brief Assessment

AI Seaports Supply Chain[41] focuses on liability law and data protection in seaport/supply chain contexts, not on HITL systems' explainability-responsibility trade-offs or GDPR/EU AI Act gaps regarding HITL requirements.

4. Evaluating the Trade-offs Between Explainability and Security in AI-Powered Cyber Defense

URL: [View paper](#)

Brief Assessment

Explainability Security Trade-offs[42] focuses on cybersecurity contexts and trade-offs between explainability and security in AI-powered cyber defense systems, not on HITL legal frameworks, GDPR/EU AI Act analysis, or responsibility attribution in HITL setups as examined in the original paper.

5. Navigating the Human-oversight Dilemma in AI-based Systems

URL: [View paper](#)

Brief Assessment

Human-oversight Dilemma[44] focuses on optimizing human oversight in AI decision-making using MCDA approaches and ontology frameworks, without analyzing specific legal frameworks (GDPR, EU AI Act) or the responsibility-explainability trade-off that the original paper identifies.

6. Toward meaningful transparency and accountability of AI Algorithms in public service delivery

URL: [View paper](#)

Brief Assessment

AI Transparency Accountability[38] focuses on algorithmic transparency and accountability in public service delivery from a policy and governance perspective, without analyzing computational formalizations of HITL setups or the specific explainability-responsibility trade-offs that emerge from different reduction types as presented in the original paper.

7. Administrative Liability for Damages Caused by Artificial Intelligence Systems in Public Services: An Analytical Study in Light of the Principles of Legality and Transparency

URL: [View paper](#)

Brief Assessment

AI Administrative Liability[43] focuses on administrative liability in public services within specific jurisdictions (France, US, Egypt, Oman), examining supervisory negligence and algorithmic mismanagement. It does not address the HITL responsibility-explainability trade-off or computational reductions framework that form the core novelty of the original paper's analysis.

8. Human-Centered Machine Learning

URL: [View paper](#)

Brief Assessment

Human-Centered Machine Learning[36] appears to focus on general human-centered ML principles rather than the specific legal analysis of GDPR/EU AI Act frameworks and the explainability-responsibility trade-off in HITL systems that the original paper examines.

9. Human-in-the-Loop AI Engineering: Enhancing Collaboration Between Developers and End Users

URL: [View paper](#)

Brief Assessment

Human-in-the-Loop AI Engineering[37] focuses on inverse accountability and transparency in AI systems generally, not on the specific legal frameworks (GDPR, EU AI Act) or the inherent trade-off between explainability and responsibility attribution in HITL setups that the original paper analyzes.

10. Percentages and reasons: AI explainability and ultimate human responsibility within the medical field

URL: [View paper](#)

Brief Assessment

AI Medical Responsibility[45] focuses on physician responsibility and explainability in medical AI contexts, not on HITL legal frameworks (GDPR/EU AI Act) or the responsibility-explainability trade-off in HITL system design that the original paper analyzes.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Formalising Human-in-the-Loop: Computational Reductions, Failure Modes, and Legal-Moral Responsibility [View paper](#)
- [1] Unknown-Unknowns in Human-Centric AI Systems [View paper](#)
- [2] Unveiling high-dimensional backstage: A survey for reliable visual analytics with dimensionality reduction [View paper](#)
- [3] LLM-based Interactive Code Generation: Empirical Evaluation [View paper](#)
- [4] Modern energy resilience studies with artificial intelligence for energy transitions [View paper](#)
- [5] Relevance-Driven Decision Making for Safer and More Efficient Human Robot Collaboration [View paper](#)
- [6] Smart computational exploration of stochastic gene regulatory network models using human-in-the-loop semi-supervised learning [View paper](#)
- [7] Tractable Agreement Protocols [View paper](#)
- [8] Reduct-based result set fusion for relevance feedback in CBIR [View paper](#)
- [9] DendroTweaks: An interactive approach for unraveling dendritic dynamics [View paper](#)
- [10] Human-in-the-loop optimization of transcranial electrical stimulation at the point of care: a computational perspective [View paper](#)
- [11] Can Humans Be out of the Loop? [View paper](#)
- [12] Using Formal Frameworks to Reduce Mode Confusion in Cyber-Physical Human-in-the-Loop Systems [View paper](#)
- [13] Adaptive temporal compression for reduction of computational complexity in human behavior recognition. [View paper](#)
- [14] Fault Detection and Human Intervention in Vehicle Platooning: A Multi-Model Framework [View paper](#)
- [15] Deep learning of individual aesthetics [View paper](#)
- [16] Towards Interactive Pattern Search in Massive Graphs [View paper](#)
- [17] Off-Policy Evaluation and Learning from Logged Bandit Feedback: Error Reduction via Surrogate Policy [View paper](#)
- [18] RRHF: Rank Responses to Align Language Models with Human Feedback without tears [View paper](#)
- [19] PoeSpin: A Human-AI Dance to Poetry System for Movement-Based Verse Generation [View paper](#)
- [20] Formal Notations and Terminology for Users' Feedback and Its Specialization for Interactive Fault Localization [View paper](#)
- [21] GAMES: A dynamic model development workflow for rigorous characterization of synthetic genetic systems [View paper](#)
- [22] Linear basis-function t-SNE for fast nonlinear dimensionality reduction [View paper](#)
- [23] Interactive Code Generation via Test-Driven User-Intent Formalization [View paper](#)
- [24] The AWGN BC with MAC Feedback: A Reduction to Noiseless Feedback via Interaction [View paper](#)
- [25] AI as Interpretive Aid in Qur'anic Stylistics: Ethical Foundations for Digital Hermeneutics [View paper](#)
- [26] Human-in-the-loop active learning for goal-oriented molecule generation [View paper](#)
- [27] Framework for Predicting Customer Sentiment Aware Queries and Results in Search Using Oracle and Machine Learning [View paper](#)
- [28] VOICE: Visual Oracle for Interaction, Conversation, and Explanation [View paper](#)
- [29] Addressing the data bottleneck in medical deep learning models using a human-in-the-loop machine learning approach [View paper](#)
- [30] Modeling Variation in Human Feedback with User Inputs: An Exploratory Methodology [View paper](#)
- [31] You are the only possible oracle: Effective test selection for end users of interactive machine learning systems [View paper](#)
- [32] Leveraging Oracle Digital Assistant (ODA) to Automate ERP Transactions and Improve User Productivity [View paper](#)
- [33] Improved Inference of Human Intent by Combining Plan Recognition and Language Feedback [View paper](#)
- [34] Towards understanding and simplifying human-in-the-loop machine learning [View paper](#)
- [35] Oracle or Teacher? A Systematic Overview of Research on Interactive Labeling for Machine Learning [View paper](#)
- [36] Human-Centered Machine Learning [View paper](#)
- [37] Human-in-the-Loop AI Engineering: Enhancing Collaboration Between Developers and End Users [View paper](#)
- [38] Toward meaningful transparency and accountability of AI Algorithms in public service delivery [View paper](#)
- [39] Requirements of high-risk AI systems: AI Act. Article 14. Human oversight [View paper](#)
- [40] Human-in-the-Loop Robotics: Enhancing Safety and Adaptability through Interactive AI Systems [View paper](#)
- [41] Autonomous AI, smart seaports, and supply chain management: Challenges and Risks [View paper](#)
- [42] Evaluating the Trade-offs Between Explainability and Security in AI-Powered Cyber Defense [View paper](#)
- [43] Administrative Liability for Damages Caused by Artificial Intelligence Systems in Public Services: An Analytical Study in Light of the Principles of Legality and Transparency [View paper](#)
- [44] Navigating the Human-oversight Dilemma in AI-based Systems [View paper](#)
- [45] Percentages and reasons: AI explainability and ultimate human responsibility within the medical field [View paper](#)
- [46] Understanding choice independence and error types in human-ai collaboration [View paper](#)
- [47] Reliability Assurance for AI Systems [View paper](#)
- [48] Human-in-the-loop Techniques in Machine Learning. [View paper](#)
- [49] Human information interaction, artificial intelligence, and errors [View paper](#)
- [50] Scientific Knowledge Graph Construction Needs an AI-Mediated, Scientist-in-the-Loop Workflow (A Blue Sky Paper) [View paper](#)
- [51] Collaborative automation in factories of the future: review and survey [View paper](#)
- [52] Context-Awareness and Interpretability of Rare Occurrences for Discovery and Formalization of Critical Failure Modes [View paper](#)
- [53] Smart Biofloc Systems: Leveraging Artificial Intelligence (AI) and Internet of Things (IoT) for Sustainable Aquaculture Practices [View paper](#)

- [54] Architecting Human-AI Systems for Effective Collaboration and Oversight: Making Sense of Human/AI in/on/Over/Under/Along the Loop [View paper](#)
- [55] Design and Application of a C++ Compiler Error Solution Query Platform Integrating Large Language Models and Human-in-the-Loop Support [View paper](#)