Paper ML36

Risk-Driven AI Oversight in Healthcare - A Tiered Framework Using ISO, ICH Q9, & FMEA

Laxmiraju Kandikatla, Maxis AI, Edison, NJ, United States
Rajesh Hagalwadi, Maxis AI, Edison, NJ, United States

ABSTRACT

Artificial intelligence (AI) is transforming healthcare and clinical research, offering opportunities to enhance efficiency, decision-making, and innovation. However, its safe adoption requires governance practices that adapt to the level of risk. This paper introduces a structured, risk-driven framework for AI oversight. The approach begins with a clear definition of each system's purpose and context, followed by risk assessment aligned with ISO 31000:2018, ISO 23894:2023, and ICH Q9. We employ Failure Modes and Effects Analysis (FMEA) to quantify risks through severity, occurrence, and detection, generating Risk Priority Numbers (RPNs) that inform targeted mitigation strategies. Oversight is then scaled proportionately to the risk and model complexity: Human-in-the-Loop for high-risk applications, Human-on-the-Loop for medium risk, and Human-in-Command for low risk. This framework is both rigorous and pragmatic, preserving human autonomy while strengthening accountability and trust. By providing a clear and scalable method, it equips healthcare stakeholders with practical tools to implement AI responsibly.

INTRODUCTION

Artificial Intelligence (AI) systems are playing an increasingly influential role across various sectors. Abilities once thought to be uniquely human—like intuition, moral judgment, and inventive thinking—are now being rivaled by machines capable not only of mimicking these traits but, in many specialized fields, surpassing them (Fui-Hoon Nah et al., 2023; Mikalef et al., 2022). As technological progress accelerates beyond the capacity of most individuals to keep up, artificial intelligence is increasingly taking center stage in reshaping how we define labor, efficiency, and even meaning itself. In this emerging landscape, AI is frequently entrusted with organizational structures, role assignment, and productivity fine-tuning (Budhwar et al., 2022). Accordingly, some studies insist that "developing machine morality is crucial, as we see more autonomous machines integrated into our daily lives" (Chu & Liu, 2023, also, Jiang et al., 2021). This is even more relevant in light of the argument maintaining that "while AI can improve efficiency, it may also reduce critical engagement, particularly in routine or lower-stakes tasks in which users simply rely on AI" (Schmidt et al., 2020; Vasconcelos et al., 2023).

The interaction between humans and machines reflects deeper existential concerns about power relations and human agency in the problem-solving process, especially given that inadequate oversight or failure to identify model errors can lead to serious and potentially irreversible harm. For example, healthcare and clinical research are particularly delicate areas due to their direct impact on patient safety, making careful implementation essential (Elgin & Elgin, 2024). Finance is highly sensitive as it directly impacts individual wealth, market stability, and fraud risks, requiring robust governance and risk controls (Moura et al., 2025).

All of this points to the need for robust governance, which should not in itself be a major issue, since policymakers, whose engagement is pivotal in shaping the future, are quick to argue that their approaches and regulations are informed by a range of factors, including past experiences and risk management strategies (Law & McCall, 2024; Wirtz et al., 2021). The process is further complicated by AI systems, which rely heavily on programming languages and coded algorithms; this suggests that responsibility for their actions extends to programmers, data scientists, content providers, institutional architects, and many others, all of whom contribute to shaping their design, training, and deployment (Gerlich, 2025). However, since human agency entails the capacity to make autonomous and informed decisions, the implementation of AI systems should support, not replace, human decision-making; in other words, rather than focusing on how AI could or will replace humans, it is more appropriate to emphasize the degree of human oversight required (Langer et al., 2025; Sturgeon et al., 2025). This question of oversight is particularly important given the potential harm that AI development and deployment may pose to individual rights. This concern underpins the argument that ethical AI design should prioritize human agency while leveraging the efficiencies of AI-driven assistance.

This paper advances the conversation on trustworthy AI (EU, 2019) by offering a structured approach to embedding human oversight models—Human-in-Command (HIC), Human-in-the-Loop (HITL), and Human-on-the-Loop (HOTL)—into risk assessment and mitigation strategies for AI systems. The paper argues that responsible AI deployment requires more than technical robustness; it also demands context-sensitive oversight aligned with the level of risk posed by the system's decisions. We first propose a practical framework for scenario identification, in which AI use cases are classified according to their potential impact on human well-being, safety, and compliance obligations. Building on this classification, we apply a risk assessment methodology to determine where human

involvement is critical and what level of oversight is required. The paper puts forward a set of guiding principles: (1) HIC ensures governance-level accountability, where humans retain ultimate authority over system objectives, as in public policy or defense applications; (2) HITL mechanisms are crucial when real-time human judgment can prevent harm, such as in clinical decision support systems; and (3) HOTL is suited for systems where monitoring and periodic intervention can mitigate risks without slowing operations, such as financial fraud detection. In addition to these oversight models, appropriate control measures—technical, procedural, and organizational—are recommended to operationalize trust and ensure compliance. In doing so, the paper contributes a repeatable, risk-based process that organizations can adopt to ensure AI systems remain safe, ethical, and aligned with human values.

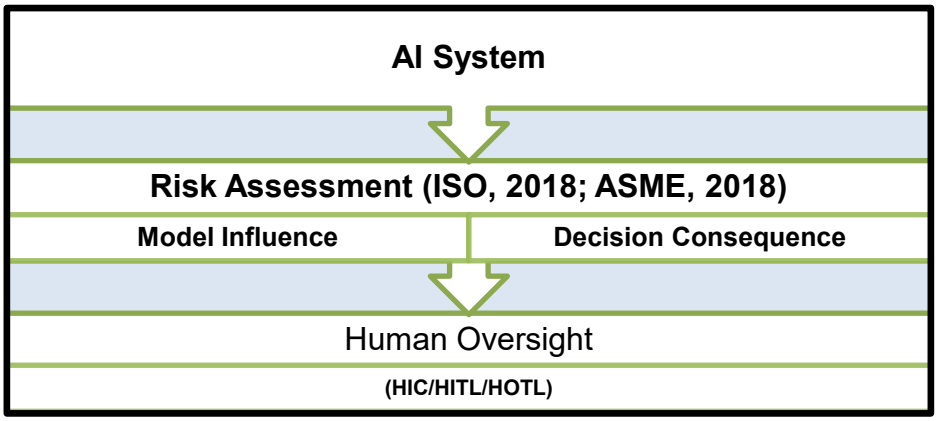

Figure 1. Risk assessment framework (Authors' proposal)

2. RISK ASSESSMENT AND AI OVERSIGHT

Effective AI risk management necessitates early and continuous communication and consultation with all relevant stakeholders, including AI developers, end-users, regulators, and domain experts (ISO, 2018). In Clinical Research Industry, embedding human oversight at this stage ensures that diverse stakeholder perspectives inform risk identification and decision-making processes, thereby enhancing transparency and fostering public trust. By integrating human judgment from the outset, organizations can more accurately anticipate the potential influence of AI outputs and the ramifications of incorrect decisions (Pandit & Rintamäki, 2024). Defining the scope, context, and risk criteria of AI deployment is crucial for determining the potential impact of AI across different applications (Figure 2). Human oversight guides the identification of high-risk scenarios where AI outputs have substantial influence on critical decisions (US FDA, 2025).

Table -1 below illustrates the alignment between ISO 31000/23894 risk management principles, ICH Q9 and E6 clinical research requirements, and FMEA activities. It highlights how risks are systematically identified, assessed, controlled, and monitored, with human oversight applied proportionally to risk. This integrated approach supports robust decision-making and regulatory compliance.

Phase	ISO 31000/23894	ICH Q9 & E6 Clinical Research	FMEA Activity
Risk Identification	Context, Scope, Hazard ID	Hazard and Risk Identification	Identify failure modes and causes
Risk Analysis	Risk Estimation (Severity, Likelihood)	Risk Estimation (Severity & Occurrence)	Rate Severity, Occurrence, Detection; Calculate RPN
Risk Evaluation	Risk Level Comparison & Prioritization	Risk Acceptability Decisions	Prioritize failure modes by RPN

Risk Treatment/Control	Selection & Implementation of Controls	Control Implementation & Monitoring	Design and implement corrective/preventive actions
	Human Oversight Mechanisms based on the Risk		
Communication & Reporting	Stakeholder Communication, Documentation	Documentation & Regulatory Reporting	Document and communicate FMEA results and actions
Monitoring & Review	Monitor Risk Residuals & Effectiveness	Ongoing Risk Review & Quality Monitoring	Reassess and update FMEA with new data

Table 1: Risk Assessment Mapping (ISO 31000, ISO 23894, ICH Q9, ICH E6, and FMEA)

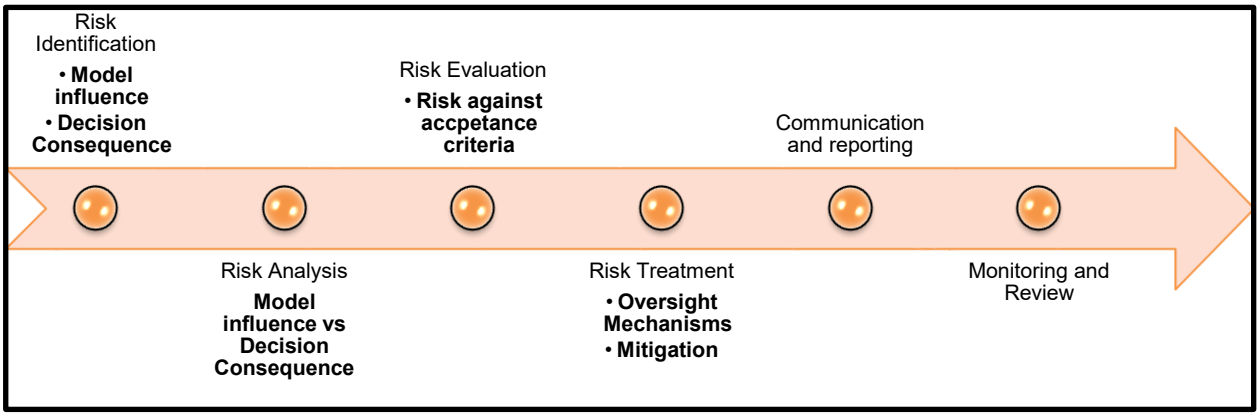

Figure 2. Risk assessment process mapping (adapted from ISO, 2018, and ASME, 2018)

2.1. RISK IDENTIFICATION

AI system risk should be conceptualized not as an intrinsic attribute of the system itself, but as a risk that arises from deploying the system to address a specific question or decision-making context. As already suggested elsewhere (ASME, 2018; DiFrancesco, 2025; US FDA, 2025), this risk can be characterized along two principal dimensions: System influence and decision consequence. System influence pertains to the relative weight attributed to the AI system's outputs compared to other sources of evidence, such as clinical studies, operational data, expert judgment, or bench testing. Decision consequences address the potential impact of an erroneous decision based on AI output. A systematic evaluation of decision consequences aligns directly with FMEA principles by considering three elements: severity of potential harm, probability of occurrence, and detectability of an error prior to adverse outcomes. These three factors are combined to derive a Risk Priority Number (RPN), enabling consistent categorization of decision consequences and supporting risk-based determination of appropriate controls, validation rigor, and human oversight mechanisms (Table 3). Assessing risk along these dimensions provides a structured framework that minimizes the risk of under- or overestimation.

Dimension	Score	Level	AI-Specific Description
Severity	1	Low	Incorrect AI model or agent decision has negligible impact and does not affect patient safety, data integrity, or regulatory compliance; impact is easily reversible.
	2	Medium	Incorrect AI decisions may cause moderate operational disruption, limited GxP impact, or require manual review or correction.
	3	High	Incorrect AI decision may result in significant harm, including patient safety risk, compromised data integrity, or regulatory non-compliance.
	1	Low	Incorrect AI decisions are unlikely to occur under intended conditions of use, given model performance, data quality, and controls.

Probability	2	Medium	Incorrect AI decisions may occur occasionally due to factors such as data variability, model limitations, or context drift.
	3	High	Incorrect AI decision is likely or frequent, driven by known model weaknesses, unstable data inputs, or operational changes.
Detectability	1	Low	High likelihood that an incorrect AI decision will be detected prior to downstream impact
	2	Medium	Incorrect AI decisions may be detected, but detection is inconsistent or dependent on manual intervention.
	3	High	Low likelihood that an incorrect AI decision will be detected before causing downstream impact due to limited monitoring

Table 2: Severity, probability and detectability categorization

Risk Priority Number (RPN) = Severity x Probability x Detectability

RPN Range	Risk Priority
1 – 6	Low
7 - 14	Medium
15 - 27	High

. Table 3: Risk Priority number range

2.2. RISK ANALYSIS

Risk Analysis for AI systems involve assessing both model risk and decision consequence to provide a comprehensive understanding of potential adverse outcomes. Model risk is characterized by the degree of influence the AI system exerts on decision-making and the potential consequences of erroneous outputs (DiFrancesco, 2025; US FDA, 2025). For instance, AI systems with high influences where outputs form a critical basis for decisions such as clinical treatment plans or financial approvals carry elevated risk compared to systems where AI insights constitute only one factor among many. Decision consequence refers to the intended outcomes and impact the system's deployment aims to achieve, which can vary from low to high severity depending on the application context. The interplay of these factors defines risk levels ranging from low, low-medium, medium, medium-high, to high (Figure 3). Higher risk levels correspond to scenarios where both the influence of the AI system is significant and the consequences of wrong decisions are severe, necessitating more stringent risk mitigation and oversight. This framework supports targeted governance approaches that consider both the systemic role of AI and the contextual consequences of its use, enabling organizations to better prioritize resources and implement safeguards mechanisms effectively.

	Decision Consequence			
	Risk Matrix	**Low**	**Medium**	**High**
Model Influence	**High**	Medium	High	High
	Medium	Low	Medium	High
	Low	Low	Low	Medium

Figure 3. Risk assessment matrix (adapted from ASME, 2018)

2.3. RISK EVALUATION

Risk evaluation involves comparing analyzed risks against predefined acceptance criteria to prioritize AI systems into low, medium, or high-risk categories (ISO, 2018). Where risks need to be reduced to acceptable thresholds for safer deployment and operation, risk treatment and control measures are selected and implemented with a particular focus on proportionate human oversight mechanisms, including Human-in-Command (HIC), Human-in-the-Loop (HITL), or Human-on-the-Loop (HOTL), selected according to the assessed risk level

2.4. RISK TREATMENT

Following risk characterization, the level of human oversight should be mapped proportionally to the identified risk category. We are proposing below oversight mechanism based on the risk matrix (Figure 4). At the highest levels of risk, particularly within safety- or life-critical contexts, a HIC oversight is warranted. Humans retain primary authority over decision-making, and the AI system functions strictly as a decision-support tool rather than an autonomous actor. For medium to medium-high risk settings, a HITL framework is more suitable, ensuring that a human decision maker validates or approves the AI system's outputs before actions are executed. For low- to medium-risk applications, risks may be appropriately managed under a HOTL framework, whereby the AI system operates autonomously but remains under continuous human supervision, with intervention triggered in response to anomalies (EU, 2019; also, Chiodo et al., 2025; Fabiano, 2025; Holzinger et al., 2025; Tomić & Štimac, 2025).

This risk-to-oversight mapping is relevant across multiple sectors. In finance, fraud detection systems may function acceptably with HOTL monitoring, whereas credit approval decisions typically require HITL review. In manufacturing, predictive maintenance systems may operate autonomously with HOTL oversight, while production shutdown decisions necessitate HITL or HIC control due to their operational consequences. In autonomous vehicles, everyday navigation may be automated, but safety-critical maneuvers such as collision avoidance require HITL confirmation or strict HIC authority. Aligning human intensity with the severity, probability, and detectability of system-related risks ensures that AI deployments achieve both efficiency and safety across diverse domains.

When AI System Risk = High and Model Influence = High and Decision Consequence = High, the system SHALL be treated as Critical for governance purposes and require Human-in-Command (HIC) oversight.

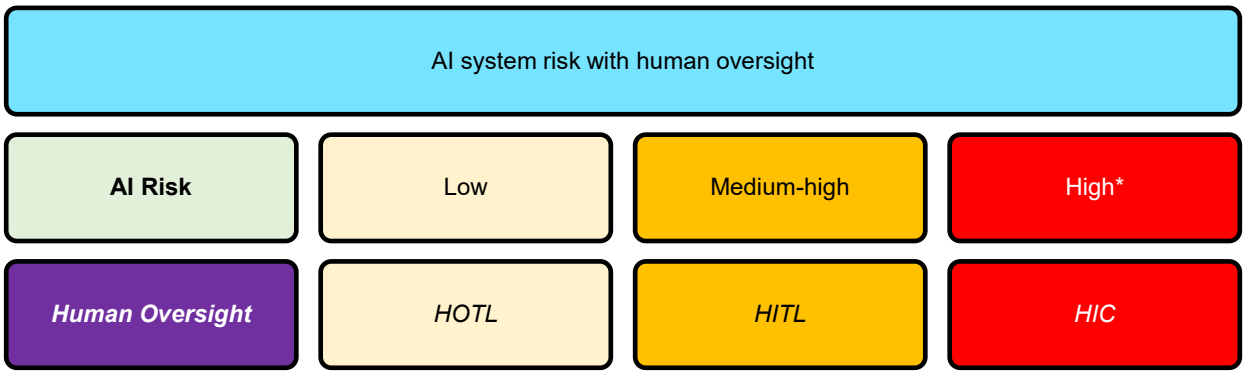

Figure 4. Human oversight mechanisms based on risk (Authors' proposal)

2.5. COMMUNICATION AND REPORTING

Communication and reporting constitute critical activity throughout the risk management process. Risk assumptions, assessment methodologies, metric results, risk categorizations, and implemented control measures are formally documented and communicated to relevant stakeholders, including developers, quality and compliance teams, clinical users, and regulatory authorities. This documentation supports traceability, auditability, and regulatory transparency, enabling informed decision-making and accountability across the AI lifecycle. (ISO, 2018; ISO/IEC, 2023).

2.6. RISK MONITORING AND REVIEW

The scope of risk monitoring and review is to ensure that mitigation strategies are effectively implemented and that AI systems remain reliable over time (Table 1). Oversight mechanisms validate AI outputs, embed fail-safes for high-consequence decisions, and establish governance protocols to clarify accountability. Continuous review, recording, and reporting enable dynamic adjustments of oversight intensity as risk assessments evolve. Across industries, this approach ensures that AI systems operate safely, human judgment complements automation appropriately, and accountability is maintained, balancing efficiency with ethical and operational responsibility (ISO, 2018; also, Madiega, 2021; Pandit & Rintamäki, 2024). Therefore, it is recommended that both model influence and decision consequence be systematically evaluated to determine the appropriate level of oversight for AI systemsContinuation of body – after source code.

ISO 31000 Step	AI-Specific Concept	Description
Risk Identification	1. Model Influence 2. Decision Consequence	1. Extent to which AI output shapes outcomes (high influence = central role in decision) and identifies potential points of failure or misuse 2. Evaluate the potential impact of an incorrect AI output by considering severity of harm, probability of occurrence, and detectability of the error before it causes harm
Risk Analysis	Combined Risk (Influence & Consequence)	Integrate model influence and decision consequence to assign a risk tier (Low, Medium, Medium-High, Critical)
Risk Treatment	Oversight Mechanism	Implement human oversight based on risk tier
Monitoring & Review	Dynamic Oversight	Continuously monitor AI performance and context of use, adjusting the level or mode of oversight as the system or associated risks evolve

Table 4. Mapping of risk evaluation with human oversight (Authors' proposal)

3. ILLUSTRATIVE RISK ASSESSMENTS INFORMING HUMAN OVERSIGHT

To illustrate the relationship between model influence, decision consequence, and the corresponding level of human oversight, three representative scenarios are examined across the domains of finance, education, and healthcare. Each case demonstrates how the same foundational AI principles can lead to varying levels of risk, depending on the system's purpose, context of use, impact, and consequences. The ASME V&V 40-2018 (ASME, 2018) framework provides a structured basis for assessing model risk by combining decision consequences (the potential impact if the model fails) with model influence (the extent to which the model informs or drives decisions). These are aligned with human oversight mechanisms (EU, 2019). Each case therefore illustrates the application of distinct human oversight models—Human-in-Command (HIC), Human-in-the-Loop (HITL), and Human-on-the-Loop (HOTL)—as mechanisms for maintaining accountability, fairness, and regulatory compliance in AI-enabled decision-making contexts. As can be observed, model risk is distinct from technical flaws, it pertains to the potential for harm when decisions are made based on flawed or misleading outputs (Pandit & Rintamäki, 2024). Human oversight, through mechanisms like HIC, HITL, and HOTL, helps mitigate such risks by addressing challenges such as opacity, bias, and lifecycle limitations while promoting fairness, safety, and compliance (EU, 2019). Regardless of the nature of the risk, human oversight helps detect and correct errors, promote fairness, maintain trust, and ensure that AI systems align with regulatory requirements, operational goals, and societal values. Each of the scenarios presented below showcase the framework's potential applicability is inspired by existing analyses and case studies.

SCENARIO: AUTOMATED PATIENT SCHEDULING

Aspect	Details
Context	An AI system automates patient appointment scheduling in a clinic or hospital. While errors may lead to delays, missed appointments, or minor patient dissatisfaction, they are unlikely to cause direct harm to the patient's health.
Model Influence	**Medium.** The AI system in this scenario does not make safety-critical decisions, as it is only involved in scheduling patients' appointments.
Decision Consequence	**Low (4)**, as incorrect scheduling may inconvenience patients or staff, but it does not directly affect patient safety. RPN = S x P x D = 1 x 1 x 3 = 3 (Low)
Evaluation	Overall, the AI system risk is **Low to Medium** (Figure 4).
Human Oversight Mechanism	**Human-on-the-Loop (HOTL) (Figure 5)** Humans monitor system outputs can intervene if anomalies occur, but daily operations largely proceed autonomously.

Table 5: Example

S – Severity; P – Probability; D – Detectability; RPN – Risk Priority Number

4. CONCLUSION

The responsible implementation of high-risk AI systems requires a nuanced understanding of both their transformative potential and their inherent risks. Effective risk assessment frameworks must account for the specific context of use, the design intentions, and the emergent risks arising from AI errors, bias, or opacity. While HITL oversight can be valuable, applying it uniformly to all AI systems may be time-consuming and burdensome. Consistent with guidance from the EU AI Act, the level of human oversight should therefore be proportionate to risk, with oversight intensity determined by factors such as the severity, probability, and detectability of potential harms (European Commission, 2021). Integrating tailored human oversight strategies—whether Human-in-Command, Human-in-the-Loop, or Human-on-the-Loop—not only enhances transparency and accountability but also helps ensure that AI-driven decisions remain aligned with ethical, legal, and societal standards (Floridi et al., 2018). Ongoing adaptation of risk assessment methods, in accordance with evolving technologies, is essential to maintain trust, protect fundamental rights, and maximize the benefits of AI innovation. Ultimately, balancing opportunity and risk through robust assessment and governance enables organizations to harness AI systems safely and fairly, fostering innovation while minimizing potential harm.

REFERENCES

Algarvio, R.C., Conceição, J., Rodrigues, P.P., et al. (2025). Artificial intelligence in pharmacovigilance: A narrative review and practical experience with an expert-defined Bayesian network tool. *International Journal of Clinical Pharmacy*, 47(4), 932–944. https://doi.org/10.1007/s11096-025-01975-3

ASME. (2018) V&V440-Assessing credibility of computational modeling through verification and validation: Application to medical devices. https://www.asme.org/codes-standards/find-codes-standards/assessing-credibility-of-computational-modeling-through-verification-and-validation-application-to-medical-devices

Bernstein, D. (2024). Who is winning the AI arms race? *Forbes*, August 28. https://www.forbes.com/sites/drewbernstein/2024/08/28/who-is-winning-the-ai-arms-race/

Budhwar, P., Malik, A., De Silva, M.T.T., et al. (2022). Artificial intelligence: Challenges and opportunities for international HRM. *International Journal of Human Resource Management*, 33(6), 1065–1097. https://doi.org/10.1080/09585192.2022.2035161

Chiodo, M., Müller, D., Siewert, P., et al. (2025). Formalising Human-in-the-Loop: Computational reductions, failure modes, and legal-moral responsibility. *arXiv*. https://doi.org/10.48550/arXiv.2505.10426

Chu, Y., & Liu, P. (2023). Machines and humans in sacrificial moral dilemmas: Required similarly but judged differently? *Cognition*, 239, 105575. https://doi.org/10.1016/j.cognition.2023.105575

DiFrancesco, J. (2025). The FDA's draft guidance, use of AI in regulatory decision-making for drug & biological products. *Avancer*, January 8. https://avancer.co/our-insights/articles/articles/harnessing-artificial-intelligence-in-drug-and-biological-product-development-a-strategic-overview-of-the-fdas-draft-guidance-for-stakeholders#:~:text=,the%20necessary%20depth%20of%20testing

Elgin, C.Y., & Elgin, C. (2024). Ethical implications of AI-driven clinical decision support systems on healthcare resource allocation: A qualitative study of healthcare professionals' perspectives. *BMC Medical Ethics*, 25(1), 148. https://doi.org/10.1186/s12910-024-01151-8

EU. (2019). Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

EU. (2024). EU Artificial Intelligence Act. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

Fabiano, N. (2025). AI Act and Large Language Models (LLMs): When critical issues and privacy impact require human and ethical oversight. *arXiv*. https://doi.org/10.48550/arXiv.2404.00600

Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). *AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations*. *Minds and Machines*, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Fui-Hoon Nah, F., Zheng, R., Cai, J., et al. (2023). Generative AI and ChatGPT: Applications, challenges, and AI–human collaboration. *Journal of Information Technology Case and Application Research*, 25(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1). https://doi.org/10.3390/soc15010006

Holzinger, A., Zatloukal, K., & Müller, H. (2025). Is human oversight to AI systems still possible? *New Biotechnology*, 85, 59–62. https://doi.org/10.1016/j.nbt.2024.12.003

ISO. (2018). ISO 31000. https://www.iso.org/standard/65694.html#lifecycle

Jiang, L., Hwang, J.D., Bhagavatula, C., et al. (2021). Can machines learn morality? The Delphi experiment. *arXiv*. https://doi.org/10.48550/arXiv.2110.07574

Journal for Research in Applies Science and Engineering Technology, 10(6), 1644–1650. https://doi.org/10.22214/ijraset.2022.44032

Knight, D.R.T., Aakre, C.A., Anstine, C.V., et al. (2023). Artificial intelligence for patient scheduling in the real-world health care setting: A metanarrative review. *Health Policy and Technology*, 12(4), 100824. https://doi.org/10.1016/j.hlpt.2023.100824

Langer, M., Baum, K., & Schlicker, N. (2025). Effective human oversight of AI-based systems: A signal detection perspective on the detection of inaccurate and unfair outputs. *Minds & Machines*, 35(1). https://doi.org/10.1007/s11023-024-09701-0

Law, T., & McCall, L. (2024). Artificial intelligence policymaking: An agenda for sociological research. *Socius*, 10. https://doi.org/10.1177/23780231241261596

Lim, J.I., Regillo, C.D., Sadda, S.R., et al. (2023). Artificial intelligence detection of diabetic retinopathy: Subgroup comparison of the EyeArt system with ophthalmologists' dilated examinations. *Ophthalmology Science*, 3(1), 100228. https://doi.org/10.1016/j.xops.2022.100228

Lu, X., Yang, C., Liang, L., et al. (2024). Artificial intelligence for optimizing recruitment and retention in clinical trials: A scoping review. *Journal of the American Medical Informatics Association*, 31(11), 2749–2759. https://doi.org/10.1093/jamia/ocae243

Madiega, T. (2021). Artificial Intelligence Act. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

Mikalef, P., Conboy, K., Eriksson, J., et al. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems*, 31(3), 257–268. https://doi.org/10.1080/0960085X.2022.2026621

Moura, L., Barcaui, A., & Payer, R. (2025). AI and financial fraud prevention: Mapping the trends and challenges through a bibliometric lens. *Journal of Risk and Financial Management*, 18(6), 323. https://doi.org/10.3390/jrfm18060323

Pandit, H.J., & Rintamäki, T. (2024). Towards an automated AI Act FRIA tool that can reuse GDPR's DPIA. *arXiv*. https://doi.org/10.48550/arXiv.2501.14756

Radeljić, B. (2024). AI as a new public intellectual? *SocArXiv Papers*. https://doi.org/10.17605/OSF.IO/49Y76

Schmider, J., Kumar, K., LaForest, C., et al. (2019). Innovation in pharmacovigilance: Use of artificial intelligence in adverse event case processing. *Clinical Pharmacology & Therapeutics*, 105(4), 954–961. https://doi.org/10.1002/cpt.1255

Schmidt, P., Biessmann, F., & Teubner, T. (2020). Transparency and trust in artificial intelligence systems. *Journal of Decision Systems*, 29(4), 260–278. https://doi.org/10.1080/12460125.2020.1819094

Smuha, N.A. (2021). From a "race to AI" to a "race to AI regulation:" Regulatory competition for artificial intelligence. *Law, Innovation, and Technology*, 13(1), 57–84. https://dx.doi.org/10.2139/ssrn.3501410

Smuha, N.A. (Ed.) (2025). *The Cambridge Handbook of the Law, Ethics, and Policy of Artificial Intelligence*. Cambridge: Cambridge University Press.

Sturgeon, B., Samuelson, D., Haimes, J., et al. (2025). Human agency bench: Scalable evaluation of human agency support in AI assistants. *arXiv*. https://doi.org/10.48550/arXiv.2509.08494

Tirumanadham, N.S.K.M.K., Thaiyalnayaki, S., & Ganesan, V. (2025). Accurate and explainable AI in student performance prediction using e-learning classification. *International Conference on Next Generation Information System Engineering (NGISE)*, Ghaziabad, India, 2025, 1–7. https://doi.org/10.1109/NGISE64126.2025.11085384

Tomić, S., & Štimac, V. (2025). Pinning down an octopus: Towards an operational definition of AI systems in the EU AI Act. *Journal of European Public Policy*. https://doi.org/10.1080/13501763.2025.2534648

US FDA. (2025). *Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products*. https://www.fda.gov/media/168399/download

Wirtz, B.W., Langer, P.F., & Fenner, C. (2021). Artificial intelligence in the public sector: A research agenda. *International Journal of Public Administration*, 44(13), 1103–1128. https://doi.org/10.1080/01900692.2021.1947319

CONTACT INFORMATION

Author Name: Laxmiraju Kandikatla
Company: Maxis AI
Address: 510 Thornall street 180, Edison, NJ 08837
Work Phone: NA
Email: laxmiraju.k@maxisit.com
Website: www.maxisit.com

Author Name: RAJESH HAGALWADI
Company: MAXIS AI
Address: 510 Thornall street 180, Edison, NJ 08837
Work Phone: 732 494 2005
Email: rajesh@maxisit.com
Website: www.maxisit.com