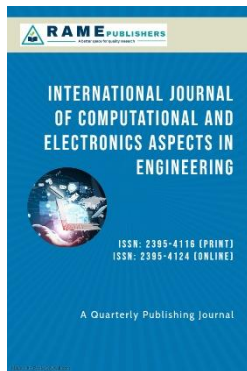


Agentic Artificial Intelligence and Machine Learning: Ethical AI and Governance Framework for the Modern Era

Sejal Waghe^{id}, Adarsh Gadpayle*, Darshana Zade, Himanshu Bangade

Department of MCA, Suryodaya College of Engineering and Technology, Nagpur, India.

*Correspondence: adarshgadpayle230@gmail.com



Article – Peer Reviewed

Received: 1 March 2026

Accepted: 30 June 2026

Published: 31 July 2026

Copyright: © 2026 RAME Publishers

This is an open access article under the CC BY 4.0 International License.



<https://creativecommons.org/licenses/by/4.0/>

Cite this article: Sejal Waghe, Adarsh Gadpayle, Darshana Zade, Himanshu Bangade, “Agentic Artificial Intelligence and Machine Learning: Ethical AI and Governance for the Modern Era”, *International Journal of Computational and Electronic Aspects in Engineering*, vol. 7, issue 3, pp. 7-14, 2026.

<https://doi.org/10.26706/ijceae.7.3.20260702>

Abstract: The current state of Agentic Artificial Intelligence which includes AI systems results from two essential technological advancements: machine learning development and large language model technology creation. Agentic AI systems use their system operations to produce decisions which they derive from their environmental analysis. The systems enable users to complete their tasks because they deliver external resources while users maintain their capability to work without needing to operate the system. Organizations which use Agentic AI across healthcare, finance, manufacturing, and digital governance will experience substantial benefits, but they must handle multiple upcoming obstacles. The system operates autonomously which results in two distinct system errors that create unexpected system behavior. Existing AI system regulations fail to control original AI systems because they do not extend their authority to systems that develop intelligence through their operational processes. The research evaluates agentic AI development through its two main themes which lead to essential ethical and governance questions, while showing that existing regulatory frameworks do not effectively resolve these important problems. The paper recommends that organizations require enhanced governance frameworks which include ongoing system assessment and establishment of precise operational standards and human monitoring to protect their systems from potential risks.

Keywords: Agentic AI, Machine Learning, AI Ethics, AI Governance, Autonomous Systems, Responsible AI

1. Introduction

Research in machine learning reached fast development during the past five years because researchers discovered two new technological advances [3][4]. The system development process creates autonomous AI systems which operate as independent entities [5].

The system operates through its perpetual loop which combines planning and executing and observing results together with performance enhancement activities throughout its entire operational period. [3] The systems enable users to execute intricate activities which require only slight involvement from humans because they control digital instruments and systems while working with users. [5] Security threats to public services and healthcare and finance sectors continue to grow which includes risks to safety and fairness and control systems. [16][18] Agentic AI systems exhibit distinct features which differentiate them from traditional AI systems because they possess the ability to execute tasks that create direct implications for physical spaces [5]. The existing governance frameworks need further development because they currently lack the capacity to handle these kinds of systems. [8][9] The research study will identify main obstacles which need resolution while developing improved methods to establish ethical standards and governance systems for agentic AI technologies [7][13].

2. Historical Evolution of Agentic AI And Governance

The Historical Development of Agentic Artificial Intelligence Together with Its Governance System The field of intelligent agents emerged from the artificial intelligence studies which began in the 1950s and continued through the 1960s. [1] The first generation of AI systems operated through two main methods which included symbolic reasoning together with rule-based systems that required prior rule establishment. [1] The systems operated through human programmed instructions yet they lacked both decision-making abilities and self-directed growth [1].

Expert systems and simple reactive agents operated in environments that contained fixed and predictable boundaries which defined their usage limitations. [1] The restricted operational capabilities of the systems during this time period led to a situation where both ethical issues and governance matters received minimal attention. [6] Machine learning technologies were developed during the 1960s which created a new fundamental change that shaped AI research for the next two decades. [2] The systems introduced new capabilities because their designers' developed systems which enabled them to study patterns through data analysis. [2] The developers created reinforcement learning systems which allowed agents to enhance their capabilities through environmental feedback mechanisms. [2] The systems displayed superior intelligence capabilities yet their usage remained restricted to laboratory and simulation environments which prevented any substantial effect on actual situations. [2] The current stage of agentic AI development has advanced through research progress which centers on both extensive neural network systems and advanced computational framework design. [3][5] The modern AI agents use large language models to operate their systems. [4]

3. Literature Review

The study conducted by Rafflesia Khan, Declan Joyce and Mansura Habiba shows their research work on "AGENTS SAFE: A Unified Framework for Ethical Assurance and Governance in Agentic AI" which aims to create a complete framework that regulates and verifies ethical standards for agentic AI systems from their inception to their operational period while securing their operational functions through safety measures and accountability systems and their ethical protection systems. [1] The research study conducted by Satyadhar Joshi in September 2025 examines AI Governance in the Era of Agentic Generative AI and AGI through two main objectives which include studying the governance risks and policy gaps and regulatory challenges that are created by agentic generative AI and AGI while developing governance frameworks that enable effective control of highly autonomous AI systems. [2] The research study by Theodore Roberts and Bahram Zarrin investigates the research study which aims to discover how AI practitioners use ethical values to create real-world agentic AI systems while determining which ethical principles become actual design and governance choices.

The International Journal of Human-Information Technology launched its Journal on Ethical and Governance Challenges of Agentic AI in September 2025 to investigate the ethical and accountability and governance issues that autonomous agentic AI systems create while highlighting the need for new governance methods that extend beyond existing AI ethical frameworks.

The research work of Mohamad Abou Ali and Fadi Dornaika in October 2025 presents an extensive study about agentic AI systems which examines their architectural designs and practical applications and anticipated future development areas. In Oct 2025 Authors Mohamad Abou Ali & Fadi Dornaika did research on "Agentic AI: A Comprehensive Survey of Architectures, Applications and Future Directions" and their Aim is to provide a comprehensive survey of agentic AI architectures, applications and future research directions, including technical foundations, real-world use cases and emerging ethical and governance concerns. [4]

4. The Need for Ethical and Governance Frameworks

The operational needs of agentic AI systems require two fundamental components which enable their technology to function independently of human intervention. [5] The system's independent functioning creates problems because it becomes impossible to determine who is accountable for failures. [14] The systems face another issue which arises from their ability to learn new functions continuously which results in unpredictable system performance. [2] The system has potential to generate two types of results which include safety threats and unexpected results. [16] The system processes extensive data inputs together with external elements which creates security vulnerabilities and privacy violations and bias identification problems. [19] The systems maintain their operational speed which makes it difficult for humans to achieve

ongoing system monitoring. [8] The system creates a risk that people will begin to rely too much on the automatic decision-making process. [11] Organizations require ongoing governance management because they must address multiple operational issues. [13] Organizations need to establish governance structures which will operate throughout the system. [13] The system needs to establish its core ethical values which include fairness and accountability and transparency from the beginning of its development. [6][7]

5. Methodological Approach

The researchers conducted the study using qualitative research methods that involved analysis of previous research about agentic AI and its ethical principles and governance practices. [9][10] The study integrates multiple research perspectives to analyze the existing research problems. [9] The NIST AI Risk Management Framework and the EU AI Act provide a framework comparison which shows their inability to manage advanced AI technologies. [8][9] The study uses real examples from healthcare and finance to solve actual problems which occur in the real world. [13][15]

5.1. Architectural Design and Governance Models

Developers of the governance-aware agentic AI system must integrate safety and ethical standards into their system designs starting from their initial development activities. [13] The system maintains its ethical alignment status throughout its entire operational lifetime. [13] The system architecture consists of four main components which include perception and reasoning and action and oversight components. [11] The perception layer collects data input from multiple sources which include data streams and sensors and user interactions. [11] The reasoning layer uses the information to find suitable actions which it identifies through its defined goals and restrictions and its understanding of the situation. [3] The action layer implements decisions through its interactions with outside systems which include executing code and using tools and interacting with users. [3] The oversight layer tracks all system operations while it checks for danger and maintains safe operational conditions and ethical operational standards. [13]

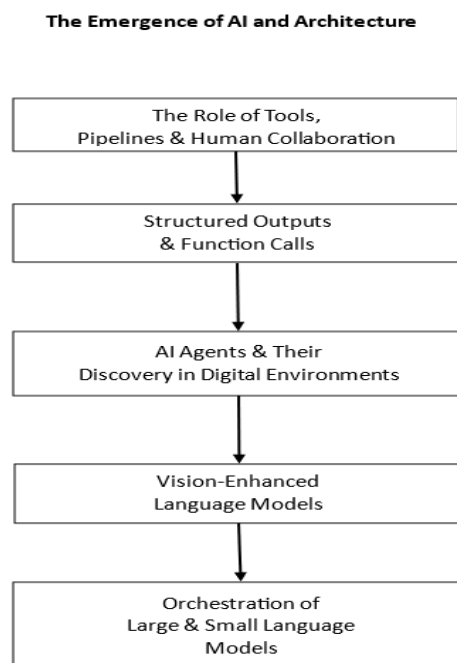


Figure 1 Agent-based AI Architecture: [2]

The decision-making process uses ethical standards which include fairness and accountability and privacy, and safety as its foundation. [6] The system conducts an ethical evaluation to assess its decisions before it takes any operational steps. [13] Governance mechanisms operate continuously to detect standard practice violations while they protect against new security threats. [8] The system implements operational restrictions which include reduced execution speed and decision escalation to human supervisors when it detects unsafe or non-compliant activities. [14]

5.2. Findings and Case Studies

The analysis of real-world use cases shows that agentic AI systems provide operational efficiency benefits for decision-making needs yet their implementation creates greater institutional hazards for organizations. The AI systems used in healthcare assist doctors with their diagnostic and treatment planning tasks yet these systems introduce major accountability challenges which arise when errors or incorrect decisions occur [16]. Autonomous trading systems in the financial industry assist market operations but they create the potential for financial markets to experience sudden breakdowns and through illegal trading activities. The implementation of agentic AI technology enables digital governance to deliver public services more effectively yet the system currently lacks adequate governance which results in existing biases to escalate and create unjust outcomes [10][18].

The two main difficulties present in these areas include: (1) organizations lack the ability to track operations constantly and (2) organizations fail to establish who should handle particular tasks. Systems that integrate strong human oversight together with clear record-keeping and explicit autonomous action limits establish higher safety standards while increasing reliability and trustworthiness. The research findings demonstrate that organizations need governance frameworks which provide active monitoring functions for their agentic AI systems instead of depending on fixed guidelines and single evaluation processes.

Human-in-the-loop Architecture:

People need to work together with machines because this system requires their combined efforts to accomplish its goals. The system uses automated decision-making which needs human judgment to handle its output through three different functions which can be used to guide its output or verify its results or create new operations.

a) Core Components

The AI model works as an automated system which makes predictions through its capability to classify data and extract insights from incoming information. The human actor represents a domain expert or end user who gives feedback together with validation and correction work and final decision-making. The interface layer lets people use the AI system through its dashboard functions and prompt systems and alert features. The human feedback system collects human input to enhance system output while developing better performance for upcoming tasks. The system uses human feedback to create new training content which will help the model grow and improve its performance throughout time.

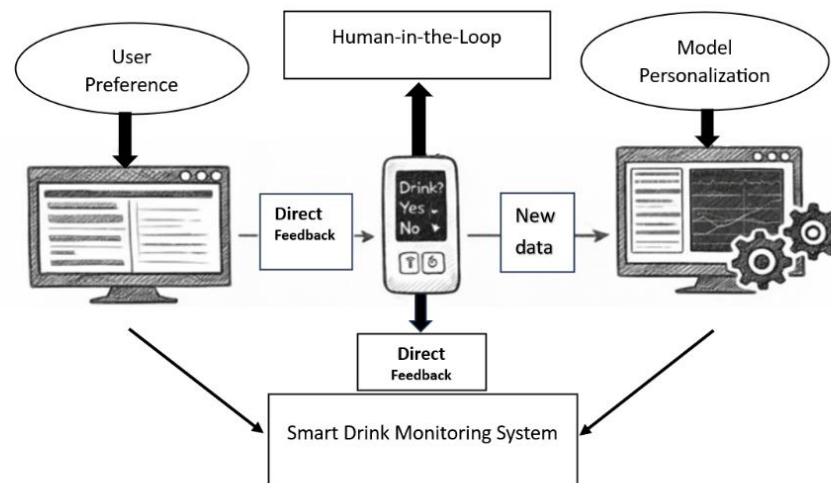


Figure 2 Human-in-the-loop Architecture: [10,11]

b) Key Characteristics

The system requires continuous human monitoring which verifies AI outputs throughout the entire process. The human error reduction process prevents both unsafe and incorrect decision-making through its active human intervention. The system presents trustworthiness and transparent operations which enable users to trust AI support throughout their work.

The system requires continuous system enhancements which will occur through ongoing user feedback. The system needs human ownership to be responsible for making important decisions. [8][14]

AI Governance Lifecycle Model:

The AI Governance Lifecycle Model provides a structured framework that controls AI systems from their initial deployment until their final shutdown while ensuring ethical operations and compliance with legal requirements and risk management and responsibility tracking.[13][17]

The main components of the core stages contain three essential elements.

The AI systems governance framework establishes the fundamental principles and ethical standards and regulatory requirements that organizations must adhere to when implementing AI technologies.[6][13]

Data Governance:

The data governance process establishes procedures for maintaining data quality while protecting privacy and security and obtaining informed consent and eliminating bias during data collection and preprocessing activities.

The model development process includes all aspects of model creation which include model design and training and validation and the creation of documentation and explainable content which must meet governance standards.

The risk assessment stage begins with risk identification which proceeds to risk mitigation of all ethical and legal and operational and security threats while ensuring regulatory compliance.

The deployment and control system manages all aspects of system deployment which includes access permission settings and version control and operational security measures.

The system auditing process continuously assesses system performance while measuring the actual bias and model drift and system security and compliance with standards through auditing procedures.

The human oversight framework establishes an accountability system which defines all roles and responsibilities and escalation paths and decision-making authority [13].

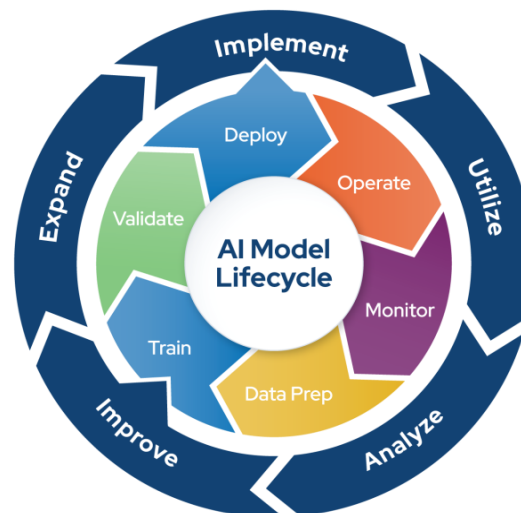


Figure 3 AI Governance Lifecycle Model: [13,15]

6. Limitations

The implementation of governance framework improvements for agentic AI systems faces multiple restrictions which prevent their complete execution. [9][14] Organizations face substantial expenses when they need to set up entire ethical frameworks and governance systems which require both technical infrastructure and ongoing monitoring capacities and

employee supervision abilities. [13] The existing regulatory framework which varies across different regions and jurisdictional boundaries creates difficulties in developing governance standards which can be applied worldwide. [9] Organizations face major difficulties when they try to measure ethical compliance because they lack standardized metrics which can evaluate context-specific concepts of fairness and accountability and transparency. [20] The implementation of strict governance regulations acts as a barrier which stops organizations from creating innovative solutions while it reduces their AI systems' operational performance and overall effectiveness. [20].

7. Conclusion

Agentic AI demonstrates advanced development progress as an AI technology solution. [5] The system requires additional responsibility capacity because it now operates with increased autonomous power. [14] The system creates unsafe conditions which require complete operational control through proper supervision. [13] System governance needs to occur during system design and implementation processes to promote safe and responsible system operations. [13] Organizations need to find a solution which enables them to develop new products while maintaining necessary security measures. [7]

8. Future Scope

Emerging research should concentrate on developing AI systems which provide enhanced safety and operational dependability. Development projects need to establish governance processes from their initial development phases. Organizations will achieve better transparency and trust through continuous monitoring combined with explainable AI systems. Human participation remains essential because it serves a critical function in major work sectors.

Global standards are required to establish effective management practices for artificial intelligence systems.

Graphical Presentation:

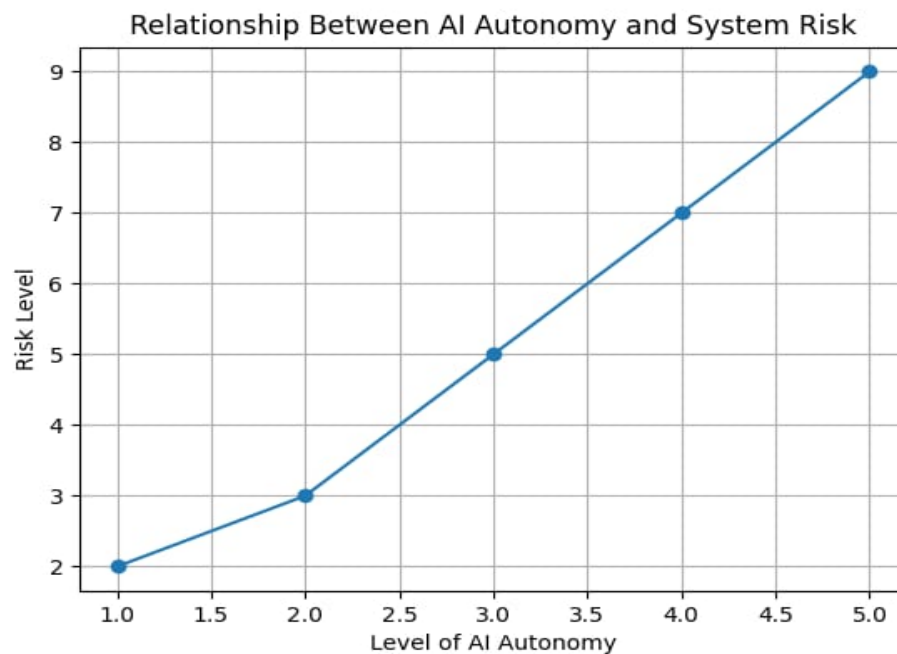


Figure 4. Relationship between AI Autonomy and System Risk

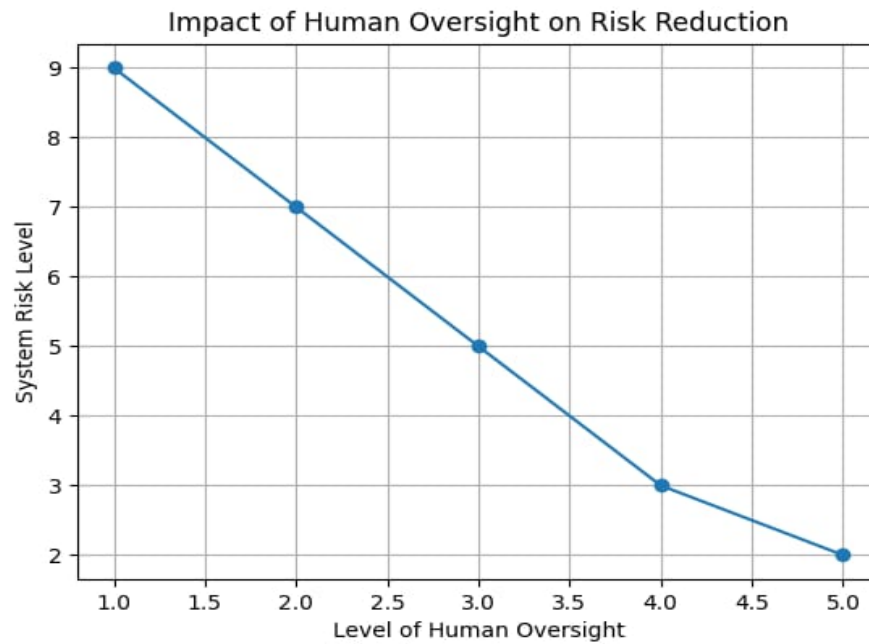


Figure 5. Impact of human oversight on risk reduction

9. Discussion

The research findings demonstrate that traditional governance methods cannot control agentic AI systems because these systems operate autonomously while they adapt and change their behavior throughout their entire existence. Agentic AI systems operate differently from traditional AI systems because they acquire knowledge through environmental interactions which enable them to create autonomous decisions and develop their capabilities through ongoing learning. The system's dynamic behavior creates a gap between existing regulatory systems and its actual operational patterns which makes it impossible to assess post-deployment through fixed ethical standards.

The research shows that organizations need to move away from their current governance approach which only reacts to problems toward creating governance systems that protect their entire processes. The governance-aware architecture shows that ethical principles should become part of system design which enables organizations to manage unexpected system behavior through ongoing system monitoring and immediate system checks. The system implements a human-in-the-loop mechanism which strengthens accountability because it requires human experts to oversee critical decisions and validate or correct them when necessary. The process reduces risks related to bias system drift and unintended outcomes while it improves user trust and system transparency.

The discussion shows that effective governance requires both technical expertise and knowledge from different fields of study. The process needs developers and policymakers together with domain specialists who will transform ethical concepts into system behavior that can be measured and enforced. The research emphasizes that governance needs to operate as a continuous lifecycle system which controls all aspects of data management model creation system implementation and ongoing system assessment.

10. Result

The research shows that agentic AI systems enhance operational efficiency and automation capabilities and decision-making processes in healthcare and financial sectors. The autonomous capabilities of these systems create obstacles that need to be solved for their organizations to establish basic accountability and transparency and effective risk management processes. The governance-aware framework solves these problems through its combination of real-time tracking systems and ethical assessment methods and human monitoring systems which enhance system protection and operational trustworthiness compared to existing static governance methods.

Uniqueness and Contribution

The research establishes its distinctiveness through its implementation of ethical principles into agentic AI systems which serve as fundamental design elements instead of operating as independent standards. The system implementation requires a dynamic governance model which adapts to system development while monitoring the system in real time. This study presents a practical implementation of its findings which tests under actual conditions while many existing studies remain theoretical. The framework achieves its strength through human-in-the-loop integration which functions as its primary element for maintaining responsibility while decreasing potential dangers in contemporary autonomous AI environments.

References

- [1] M. Wooldridge, *An Introduction to Multiagent Systems*, 2nd ed. Chichester, U.K.: Wiley, 2009.
- [2] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
- [3] S. Yao et al., “ReAct: Synergizing reasoning and acting in language models,” *arXiv preprint arXiv:2303.11366*, 2023.
- [4] OpenAI, “GPT-4 technical report,” *arXiv preprint arXiv:2303.08774*, 2023.
- [5] J. S. Park et al., “Generative agents: Interactive simulacra of human behavior,” *arXiv preprint arXiv:2304.03442*, 2023.
- [6] L. Floridi et al., “AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations,” *Minds and Machines*, vol. 28, no. 4, pp. 689–707, 2018.
- [7] A. Jobin, M. Ienca, and E. Vayena, “The global landscape of AI ethics guidelines,” *Nature Machine Intelligence*, vol. 1, pp. 389–399, 2019.
- [8] National Institute of Standards and Technology (NIST), *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1, Gaithersburg, MD, USA, 2023.
- [9] European Union, *Regulation (EU) 2024/1689 of the European Parliament and of the Council: Artificial Intelligence Act*, 2024.
- [10] OECD, *OECD Principles on Artificial Intelligence*, Paris, France: OECD, 2019.
- [11] S. Amershi et al., “Power to the people: The role of humans in interactive machine learning,” *AI Magazine*, vol. 35, no. 4, pp. 105–120, 2014.
- [12] A. Holzinger, *Interactive Machine Learning for Health Informatics*. Cham, Switzerland: Springer, 2016.
- [13] R. Khan, D. Joyce, and M. Habiba, “AGENTS SAFE: A unified framework for ethical assurance and governance in agentic AI,” *arXiv preprint*, 2025.
- [14] S. Joshi, “AI governance in the era of agentic generative AI and AGI: Frameworks, risks and policy directions,” 2025.
- [15] IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, IEEE Standards Association, 2022.
- [16] E. J. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York, NY, USA: Basic Books, 2019.
- [17] A. A. Kirilenko, A. S. Kyle, M. Samadi, and T. Tuzun, “The flash crash: The impact of high frequency trading on an electronic market,” *The Journal of Finance*, vol. 72, no. 3, pp. 967–998, 2017.
- [18] World Economic Forum, *Global AI Governance Toolkit*, Geneva, Switzerland: World Economic Forum, 2023.
- [19] F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” *arXiv preprint arXiv:1702.08608*, 2017.
- [20] L. Floridi, *The Ethics of Artificial Intelligence*. Oxford, U.K.: Oxford University Press, 2023.