Jurnal Masyarakat Informatika, 16(2), 2025
DOI:10.14710/jmasif.16.2.7708
ISSN: 2777-0648

A Critical Review of Agentic AI: Core Technologies, Applications, Ethical Implications, and Future Research Directions

Sumit Goyal*
Data Science Consultant, Global Consult, Kyndryl, India
* *Corresponding author: thesumitgoyal@gmail.com*

Abstract
Artificial intelligence (AI) is progressing toward the Agentic AI paradigm, which involves intelligent systems capable of autonomous, proactive, and goal-focused behavior through adaptive interactions with their environment. This article provides a critical review of the development of Agentic AI, examining its technological foundations, application areas, and the associated technical, ethical, and policy challenges. The review employs a narrative approach, examining primary literature from the IEEE, Scopus, and ScienceDirect databases for the period 2019–2025, using keywords such as agentic AI, multi-agent systems, human–AI collaboration, and autonomous decision systems. The findings are organized into a three-layer conceptual framework that links core technologies, such as Reinforcement Learning, Multi-Agent Systems, and Natural Language Processing, with various application domains and cross-cutting challenges. The analysis indicates that despite the significant potential of Agentic AI, gaps remain in areas such as agent interoperability, autonomy assessment metrics, and field implementation limitations. This article proposes a structured research agenda aimed at developing Agentic AI that is more transparent, trustworthy, and aligned with human values.

Keywords: *Agentic AI, Generative AI, Artificial Intelligence, Cloud Computing, Machine Learning*

1 Introduction

Artificial intelligence (AI) has evolved dramatically over recent decades from rule-based systems and expert algorithms to sophisticated models capable of flexible, autonomous, and adaptive behavior. In recent years, there has been an increase in the development of "agentic AI," which refers to AI systems characterised by intentionality, autonomy, and the capability to pursue complex, long-term objectives. As demand for real-time decision-making and autonomous functions increases, agentic AI is being integrated into socio-technical systems, resulting in changes across various industries and societal structures. The emergence of "agentic AI" signals a pivotal moment in this evolutionary arc, representing a new generation of intelligent systems that are imbued with a sense of agency, intentionality, and the capacity for independent action. Unlike traditional AI, which often relies on explicit programming and pre-defined responses, agentic AI systems are engineered to exhibit purposeful behavior, adapting dynamically to their environment and learning from ongoing interactions. These systems possess the capability to perceive, interpret, and respond to complex scenarios, enabling them to make critical decisions in real-time without the need for constant human supervision. This shift toward agentic intelligence is driven by advances in machine learning, reinforcement learning, symbolic reasoning, and collective intelligence, which collectively empower AI agents to tackle multifaceted, long-term goals. In sectors such as healthcare, finance, logistics, and

Received: 18 August 2025; Revised: 19 October 2025; Accepted: 24 October 2025; Published: 7 November 2025

autonomous vehicles, the ability for machines to reason, adapt, and act on their own behalf is unlocking unprecedented possibilities, from real-time medical diagnostics and robotic caregivers to self-driving fleets and adaptive financial advisors [1].

As technology advances, agentic AI is rapidly becoming a driving force behind the digital transformation of society, challenging conventional notions of automation and intelligence. However, this rapid evolution is accompanied by new technical, ethical, and societal challenges. The deployment of AI agents with autonomous decision-making capabilities presents significant considerations regarding accountability, transparency, safety, and the broader impacts on human collaboration and oversight [2]. This paper delves deeply into the landscape of agentic AI, offering a thorough examination of its conceptual foundations, enabling technologies, varied applications, and the open challenges that lie ahead. By synthesizing current research and practical implementations, it seeks to provide a comprehensive overview that both informs and inspires further exploration into the transformative potential and responsible stewardship of agentic artificial intelligence.

2 Background and Definition of Agentic AI

To understand the emergence of Agentic AI, it is essential to examine the theoretical underpinnings that distinguish agentic behavior from traditional AI paradigms. This section outlines the conceptual evolution and defining characteristics of Agentic AI, positioning it within the broader context of cognitive and generative intelligence. Agentic AI encompasses artificial agents that exhibit purposeful, autonomous behavior, often directed toward specific goals, whether defined by their original programming or acquired through learning and adaptation. These agents can range from relatively simple, rule-based constructs to complex systems that leverage deep learning, symbolic reasoning, and collective intelligence. Key components of Agentic AI are presented in Table 1.

Table 1 Agentic AI key components

Aspect	Description
Agentic AI	Artificial agents that exhibit purposeful, autonomous behavior directed toward specific goals, defined by programming or acquired through learning and adaptation
Range	Simple rule-based constructs to complex systems leveraging deep learning, symbolic reasoning, and collective intelligence
Agency	Capacity to analyze situations, make context-sensitive decisions, and execute actions with minimal external oversight
Autonomy	Ability to operate independently, adapting strategies based on interactions with unpredictable and evolving environments
Goal Orientation	Pursuit and prioritization of objectives involving multi-step reasoning, planning, and negotiation with other agents or humans
Characteristics	Adaptability, responsiveness, ability to manage complexity in domains where static or purely reactive approaches are inadequate

Agentic AI systems are characterized by their adaptability, responsiveness, and ability to manage complexity in domains where static or purely reactive approaches are inadequate [3].

3 Core Technologies and System Frameworks

Building upon the conceptual foundations discussed earlier, this section explores the core technologies that enable agentic behavior in modern AI systems. By integrating reinforcement learning, multi-agent coordination, natural language processing, and vision-based perception, Agentic

AI frameworks enable autonomous reasoning and continuous adaptation across diverse environments, as illustrated in Figure 1.

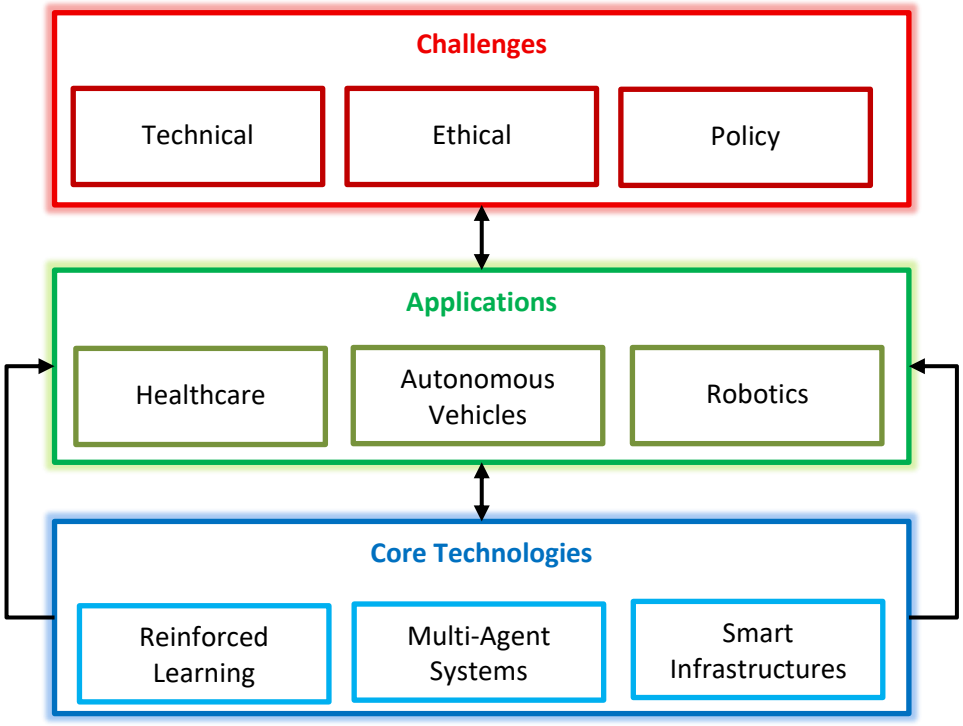

Figure 1 Conceptual framework of Agentic AI: from core technologies to ethical challenges

3.1 Core Technologies Enabling Agentic AI

Agentic AI represents an advanced approach in artificial intelligence, focusing on creating autonomous systems capable of proactive decision-making and interacting with complex environments. Its foundational technologies cover various AI subfields, each contributing distinct capabilities:

a) *Reinforcement Learning (RL)*: Agents learn to maximize cumulative rewards by interacting with their environment, employing algorithms such as Q-learning, deep Q-networks (DQN), and policy gradients. RL is crucial in Agentic AI, where systems learn optimal actions through interactions with the environment to maximize cumulative rewards. It finds applications in dynamic environments such as autonomous driving and energy management, allowing systems to adapt and learn strategies based on feedback from actions taken [4, 5].

b) *Multi-Agent Systems (MAS)*: Multiple AI agents interact, communicate, and coordinate to achieve individual or collective objectives. MAS frameworks facilitate collaboration, competition, negotiation, and emergent behaviors in complex environments. These agents are crucial for scenarios that require distributed and cooperative problem-solving. These systems involve multiple interacting agents that collaboratively work on tasks such as resource allocation and intrusion detection, providing decentralized solutions and addressing agent coordination and security issues [6, 7].

c) *Symbolic Reasoning & Automated Planning*: Agents use structured representations and logic-based methods to plan sequences of actions, enabling long-term strategic thinking and symbolic manipulation. Symbolic reasoning involves rule-based systems that facilitate logical inferences, which are crucial for tasks demanding explicit reasoning, such as knowledge representation and deductive logic [8, 9].
d) *Cognitive Architectures*: Models such as SOAR, ACT-R, and LIDA emulate aspects of human cognition, memory, perception, attention, and learning, creating AI agents that reason and adapt more naturally. These architectures model human-like cognitive processes, incorporating mechanisms such as memory and learning, which are key for tasks requiring high-level reasoning and decision-making, akin to human thinking [10].
e) *Natural Language Processing (NLP)*: NLP enables agents to process human language, understand instructions, and interact conversationally, facilitating integration into human-centric workflows. NLP empowers Agentic AI by enabling systems to comprehend and interpret human language, thereby enhancing their interaction capabilities. It is widely applied in healthcare, education, and government for tasks such as data analysis and sentiment evaluation, despite challenges like semantic nuances and data quality issues [11,12].
f) *Computer Vision and Perception*: Visual processing capabilities allow agents to interpret and *navigate* physical or digital spaces, detect objects, understand scenes, and respond to environmental cues. It can interpret visual data from the environment, which is crucial for automated systems performing tasks such as object detection and scene understanding in domains like surveillance and robotics [13, 14].
g) *Meta-Learning*: "Learning to learn" paradigms empower agentic AI systems to adapt rapidly to new tasks, generalizing knowledge and strategies across domains [15 – 17]. This learning paradigm enables AI systems to improve their learning processes based on previous experiences, promoting adaptability and efficiency in new and varied tasks [18].

Agentic AI integrates these diverse technological pillars to form systems capable of sophisticated, autonomous decision-making, addressing practical challenges and leveraging interdisciplinary collaboration for continuous enhancement [16] [19]. Such systems have promising applications across various industries, including healthcare, finance, and energy, facilitating adaptive and intelligent operations with minimal human intervention [4, 9].

3.2 Agentic AI Framework

A practical agentic AI framework serves as the backbone for orchestrating complex autonomous systems, bringing together the technological components, protocols, and design philosophies that empower agents to act intelligently and collaboratively. At its core, a robust framework encapsulates several key pillars: modularity, interoperability, autonomy, and adaptability. Within such a framework, agents are not isolated; instead, they operate as part of a distributed network, exchanging information and coordinating actions in real time. This architecture is typically layered, with foundational modules handling communication and security, while higher-level layers manage learning, reasoning, and decision-making. Open standards and APIs are fundamental, ensuring seamless integration with existing infrastructure and facilitating communication among heterogeneous agents and legacy systems. A hallmark of agentic AI frameworks is their support for flexible agent hierarchies. Individual agents can represent simple sensors or actuators, while more complex, composite agents oversee coordination, optimization, or even cross-domain orchestration. This layered

approach allows organizations to scale deployments incrementally from local, task-specific agents to global, cross-functional networks [17, 20, 21].

Adaptability is woven into the framework's fabric. By embracing plug-and-play modules and standardized interfaces, developers can rapidly prototype, test, and iterate on agent behaviors. Continuous learning mechanisms, driven by feedback loops and reinforcement learning, enable agentic AI systems to evolve in tandem with changing environments, regulations, and stakeholder needs. Security and governance are embedded from the outset. Frameworks integrate robust authentication, authorization, and monitoring tools, ensuring the integrity, transparency, and accountability of agent operations. This foundation not only safeguards against emerging threats but also builds public trust in the expanding role of agentic AI [2].

In summary, the agentic AI framework is more than just software. It is a dynamic ecosystem that enables distributed, intelligent, and resilient agents to shape the digital future. Its modularity, openness, and emphasis on continuous improvement position agentic AI as a cornerstone of tomorrow’s autonomous solutions.

3.3 Agentic AI & Cloud Computing

The symbiosis between agentic AI and cloud computing represents a pivotal advancement in the deployment and scalability of autonomous systems. Cloud infrastructure provides the computational backbone necessary to support the distributed, data-intensive nature of agentic AI, enabling agents to operate seamlessly across geographies and domains. Through cloud-native architectures, agentic AI frameworks gain access to virtually limitless storage, scalable processing power, and high-speed connectivity. This empowers agents to ingest, analyze, and act upon vast streams of data in real time, whether orchestrating logistics across continents or optimizing energy grids at the city scale. The elasticity of the cloud allows organizations to scale agent populations dynamically, adjusting resources on demand as workloads fluctuate. Moreover, cloud platforms foster interoperability by connecting heterogeneous agents, legacy systems, and third-party services through standardized APIs and integration layers. This environment not only accelerates the development and deployment of agentic solutions but also ensures consistent governance, monitoring, and lifecycle management. Centralized dashboards and analytics tools in the cloud offer unified visibility into agent operations, while continuous deployment pipelines enable rapid updates and iterative improvement [22].

Security and compliance are further strengthened when agentic AI leverages cloud best practices: encrypted data transmission, role-based access controls, and automated incident response mechanisms. Cloud providers invest heavily in meeting international regulatory standards, which helps organizations maintain trust and compliance as their agentic ecosystems expand. Crucially, cloud computing democratizes access to advanced agentic capabilities. Smaller organizations and individual developers can now harness the same powerful tools and infrastructure as global enterprises, accelerating innovation and broadening participation in the agentic AI revolution [23].

3.4 Agentic AI Tools & Implementation

Agentic AI’s transformative capabilities are realized through a diverse ecosystem of tools and implementation frameworks. At the foundation, agentic platforms leverage modular software

architectures that enable distributed intelligence, interoperability, and autonomous decision-making [24].

a) *Multi-Agent Platforms*: Frameworks such as OpenAI Gym, JADE (Java Agent Development Framework), and Ray facilitate the development, training, and deployment of agentic systems. These platforms support collaborative behavior, inter-agent communication, and robust coordination, allowing agentic AI to scale from laboratory prototypes to global applications.
b) *Edge-to-Cloud Integration*: Agentic AI often bridges edge devices with cloud infrastructure, enabling real-time sensing, decentralized processing, and adaptive coordination. Technologies such as IoT middleware, federated learning, and container orchestration, such as Kubernetes, play crucial roles in ensuring agents remain responsive and resilient even in highly volatile environments.
c) *Toolkit Ecosystem*: Developers can use specialized libraries such as Stable Baselines, Rllib, spaCy, Hugging Face Transformers, and explainable AI tools like LIME and SHAP. These toolkits simplify building agent behaviors, user interfaces, and analytics dashboards, enabling quick prototyping and continuous iteration improvement.
d) *Security and Governance Modules*: To ensure safe deployment, agentic AI implementations often include encryption protocols, sandboxing environments, and continuous monitoring tools. Audit trails, access control mechanisms, and automated compliance checks are integrated to ensure transparency and adherence to regulatory standards.

Successful implementation demands a multidisciplinary approach. Solution architects collaborate with domain experts, data scientists, and end-users to define objectives, deploy agents, and gather feedback for ongoing refinement. Pilot programs, ranging from autonomous energy management in smart grids to personalized learning assistants in education, illustrate the versatility and capacity of agentic AI for real-world impact. As organizations embrace agentic architectures, the focus shifts to scalability, maintainability, and ethical alignment. Continuous monitoring, open standards, and robust documentation are essential for sustaining trust and ensuring that agentic AI solutions remain both effective and responsible [25–29].

4 Applications and Case Studies

The versatility of Agentic AI becomes evident through its applications across multiple domains, including autonomous vehicles, intelligent healthcare systems, financial analytics, and smart infrastructure. This section examines key use cases and empirical evidence demonstrating the transformative potential and operational challenges of Agentic AI in real-world scenarios.

4.1 Applications of Agentic AI

Agentic AI's versatility is evident in its widespread adoption across various sectors, including healthcare, autonomous vehicles, robotics, finance, education, and smart or intelligent infrastructure environments.

1) Healthcare

The healthcare sector benefits substantially from agentic AI's ability to process complex data, make autonomous recommendations, and provide personalized interventions.

a) *Clinical Decision Support*: Agentic AI systems, such as IBM Watson, mine electronic health records, medical publications, and genomic data to suggest diagnoses and treatment plans tailored to individual patients. These systems help clinicians manage information overload and improve diagnostic accuracy.
b) *Remote Monitoring and Predictive Care*: Wearable devices and smart sensors powered by agentic AI continuously monitor patient vitals, detect anomalies (such as arrhythmias), and trigger early interventions. For example, AI agents can adjust insulin dosages for diabetic patients in real time.
c) *Robotic Surgery*: Surgical robots, like the da Vinci Surgical System, use agentic AI to assist surgeons, enhance precision, and adapt intraoperatively to patient-specific anatomy. Some research prototypes autonomously suture tissues or navigate complex anatomical pathways.
d) *Drug Discovery and Development*: AI agents autonomously explore chemical spaces, simulate molecular interactions, and optimize experimental designs, accelerating the identification of promising drug candidates and reducing costs [30].

2) Autonomous Vehicles

Autonomous transport systems exemplify the forefront of agentic artificial intelligence in practical application.

a) *Self-Driving Cars:* Companies such as Waymo and Tesla employ agentic AI to interpret sensor data, predict traffic behaviors, make split-second decisions, and plan safe routes. These vehicles can handle complex scenarios such as urban navigation, merging traffic, and obstacle avoidance, largely without human input.

b) *Drones and UAVs*: Agentic AI enables unmanned aerial vehicles to handle parcel delivery, such as Zipline, environmental monitoring, and disaster response. These drones autonomously develop flight plans, avoid obstacles, and adjust to real-time weather changes.

c) *Public Transit Optimization*: Agentic systems dynamically adjust transport schedules, reroute buses or trains, and optimize passenger flow based on real-time sensor data and predictive analytics, improving efficiency and minimizing delays [3,16,17].

3) Robotics

Robotic systems enhanced by agentic AI are rapidly transforming industrial automation, domestic service, and field operations.

a) *Manufacturing Automation*: Agentic robots autonomously coordinate assembly lines, manage inventory, and respond dynamically to supply chain disruptions. Collaborative robots (“cobots”) work safely alongside humans, adjusting their tasks and behavior on the fly.

b) *Service Robots*: In settings such as hospitals and hotels, agentic AI enables robots to deliver medications, disinfect rooms, or provide concierge services, adapting to new instructions and environmental changes.

c) *Search and Rescue*: Robots equipped with agentic AI explore collapsed buildings, map hazardous sites, and locate survivors after disasters such as earthquakes or fires, often communicating and collaborating as a team.

d) *Agricultural Robotics*: Smart farm robots autonomously plant, irrigate, and harvest crops, using agentic AI to optimize growth conditions and manage resources efficiently, adjusting strategies as environmental variables change [30,31].

4) Finance

Agentic AI is redefining the financial sector by automating complex, high-stakes decision-making.

a) *Algorithmic Trading*: AI agents monitor market trends, analyze financial news, detect arbitrage opportunities, and autonomously execute trades at microsecond speeds. Multi-agent trading systems interact in global markets, exhibiting emergent behaviors such as flash crashes.

b) *Portfolio Management*: Robo-advisors personalized by agentic AI construct and rebalance portfolios based on risk preferences, life goals, and market conditions, providing bespoke financial advice to individuals at scale.

c) *Fraud Detection and Prevention*: AI agents autonomously monitor transactions for suspicious activity, adapt to evolving fraud tactics, and flag or block high-risk operations in real time, minimizing losses and enhancing security.

d) *Credit Scoring*: Agentic AI analyzes diverse data sources such as transactional, behavioral, and social, that enable fairer, more comprehensive credit assessments, particularly for underserved populations [17,21, 31].

5) Education

Agentic AI is revolutionizing education by facilitating personalized and adaptive learning on a large scale.

a) *Intelligent Tutoring Systems*: AI agents assess student strengths, weaknesses, and learning styles; provide real-time feedback; and modify instruction paths to optimize learning outcomes. Examples include Carnegie Learning and Squirrel AI.

b) *Administrative Automation*: Autonomous agents streamline back-office operations, automate grading, assign resources, and manage schedules, freeing educators to focus on teaching.

c) *Student Engagement and Wellbeing*: Chatbots and virtual assistants powered by agentic AI help students navigate academic and personal challenges, providing guidance, reminders, and mental health support.

d) *Remote and Lifelong Learning*: Agentic AI enables adaptive e-learning platforms that support learners across age groups, career stages, and geographies, promoting continuous upskilling and digital inclusion [32].

6) Smart Infrastructure and Environment

Agentic AI constitutes a fundamental element of the vision for intelligent, sustainable, and resilient infrastructures.

a) *Smart Grids*: Distributed agentic AI systems autonomously manage energy generation, storage, and distribution, balancing supply and demand, integrating renewables, and rapidly responding to disruptions.

b) *Urban Mobility*: Agentic AI orchestrates traffic lights, ride-sharing, and last-mile logistics, minimizing congestion, reducing emissions, and optimizing urban flows in real time.

c) *Environmental Monitoring and Conservation*: Swarms of AI agents deploy sensors, analyze air, water, and soil quality, and trigger interventions to mitigate pollution or natural disasters. Agentic AI aids in wildlife tracking, anti-poaching patrols, and sustainable land management.

d) *Disaster Response*: Autonomous agents coordinate emergency alerts, evacuation routes, and resource allocation during crises, improving community resilience and response times [30,31].

Table 2 outlines the strengths, limitations, and key lessons associated with Agentic AI applications.

Table 2 Strengths, Limitations, and Key Lessons

Strengths	Limitations	Key Lessons
Enhances decision-making through data-driven insights	Potential for bias if trained on unrepresentative data	Continuous monitoring and validation are essential
Automates routine administrative tasks	Interpretability and explainability remain challenges	Human oversight is vital for complex decisions
Improves efficiency and reduces human error	Privacy and data security concerns	Transparent data governance frameworks must be adopted
Enables personalized treatment and patient engagement	Integration issues with legacy healthcare systems	Collaboration between clinicians and AI developers is crucial
Supports remote monitoring and applications	High implementation and maintenance costs	Scalable solutions require careful resource planning

4.2 Case Studies

a) *IBM Watson in Oncology*: IBM Watson offers oncologists autonomous decision support by analyzing patient records, current medical research, and clinical trial data. It recommends individualized treatment regimens that account for genetic markers and comorbidities. While Watson demonstrates the power of agentic AI in augmenting medical expertise, ongoing studies evaluate its integration into health systems and its real-world clinical outcomes.

b) *Waymo Autonomous Vehicles*: Waymo's self-driving vehicles rely on agentic AI to continuously interpret diverse sensor inputs, predict the intentions of other road users, and plan safe, lawful responses to a myriad of situations. Waymo vehicles adapt strategies for adverse weather, construction, or unexpected obstacles, illustrating the maturation of agentic autonomy in transportation.

c) *Zipline's Autonomous Drone Delivery*: Zipline's agentic AI-powered drones deliver essential medical supplies to remote clinics and disaster zones. The drones plan optimal routes, adjust to airspace regulations, reroute around storms, and reliably execute deliveries in challenging environments with minimal human supervision. Zipline's system has saved thousands of lives by overcoming logistical barriers.

d) *OpenAI's ChatGPT and Conversational Agents*: Conversational AI systems like OpenAI's ChatGPT demonstrate agentic characteristics by maintaining context, understanding nuanced requests, and autonomously generating human-like responses across a range of topics. These

agents facilitate customer service, mental health support, and productivity tools, raising questions about transparency, safety, and alignment with user values.

e) *AlphaGo and Competitive Gaming*: DeepMind's AlphaGo exemplifies agentic AI in competitive environments. By combining deep neural networks and Monte Carlo tree search, AlphaGo learned to outperform human champions at the ancient game of Go, developing innovative strategies and demonstrating the potential for agentic AI in decision-rich, adversarial contexts [33 – 37].

4.3 Advantages of Adopting Agentic AI

Agentic AI, characterized by autonomous decision-making, adaptability, and learning capabilities, presents a transformative opportunity for numerous sectors. Its adoption offers a multitude of distinct advantages that extend from operational efficiencies to societal impact.

a) *Enhanced Efficiency and Productivity:* Agentic AI is great at automating complex, repetitive, or urgent tasks, helping organizations boost productivity with fewer resources. By handling processes like grading, resource allocation, energy distribution, and logistics, agentic systems reduce human error, cut costs, and speed up decision-making in education, infrastructure, and industry.

b) *Personalized and Adaptive Experiences:* A key advantage of agentic AI is its ability to customize experiences to individual needs. In education, this involves real-time adaptive learning paths and targeted support, tailored to each student's strengths and weaknesses. In urban and healthcare settings, agentic AI adjusts services and interventions dynamically, resulting in better outcomes for diverse populations.

c) *Scalability and Responsiveness:* Agentic AI's distributed architecture enables smooth scaling across different regions and populations. Whether managing energy grids, tracking environmental conditions, or handling large-scale emergencies, agentic agents can operate independently, respond quickly to disruptions, and ensure continuous service even under challenging conditions of stress.

d) *Continuous Learning and Improvement:* Unlike traditional programmed systems, agentic AI agents learn from data and feedback, refining their models and behaviors over time. This ensures that deployed systems stay current with changing environments, regulations, and user expectations, fostering ongoing improvement without manual intervention.

e) *Augmented Human Capabilities:* Instead of just replacing human labor, agentic AI enhances human decision-making and creativity. In education, it allows educators to concentrate on mentorship and advanced teaching. In city management, it gives planners actionable insights and real-time analytics. Across various fields, agentic AI assists professionals by managing routine tasks, highlighting crucial information, and delivering intelligent support recommendations.

f) *Enabling Innovation and New Services*: Agentic AI opens the way for new business models and services that were previously unimaginable. Adaptive e-learning platforms support lifelong education, innovative mobility solutions transform urban dynamics, and AI-driven conservation strategies enable proactive environmental stewardship. The flexibility and autonomy create opportunities for experimentation, rapid prototyping, and cross-sector collaboration.

g) *Resilience and Sustainability*: By autonomously managing resources, anticipating disruptions, and orchestrating coordinated responses, agentic AI improves the resilience of critical infrastructure and systems. Its role in optimizing energy consumption, monitoring environmental conditions, and responding to disasters directly supports global sustainability and climate goals [20, 38 – 41].

Taken together, these advantages illustrate why agentic AI is increasingly viewed as foundational to the next wave of digital transformation. Its adoption holds the promise not only of operational gains but also of broader social progress, inclusivity, and sustainability.

5 Ethical, Technical, and Societal Implications

Agentic AI systems are evolving from controlled environments to open-world deployment, and new layers of ethical, technical, and societal complexity are emerging. This section discusses these multidimensional implications and emphasizes the importance of responsible design, explainability, and human oversight.

a) *Accountability*: Determining responsibility for decisions made by autonomous agents, especially in high-stakes contexts, remains unresolved. Legal and regulatory frameworks must evolve to address liability and transparency.
b) *Bias and Fairness*: Agentic AI systems can inadvertently perpetuate or amplify social biases present in training data or decision processes. Ongoing research is required to ensure fairness, equity, and inclusivity.
c) *Transparency & Explainability*: Many agentic AI models, especially deep learning, function as “black boxes”, which makes it hard for stakeholders to interpret or trust their decisions. Explainable AI is essential for adoption in sensitive areas and domains.
d) *Security and Adversarial Robustness*: Autonomous agents are vulnerable to hacking, spoofing, or adversarial attacks that could compromise their integrity or mission. Robust safeguards and verification mechanisms are essential.
e) *Human-AI Interaction*: As agentic AI systems increasingly operate alongside humans, designing effective collaboration protocols, user interfaces, and trust-building mechanisms is vital.
f) *Societal Impact*: Agentic AI may reshape labor markets, exacerbate digital divides, or influence democratic processes. Policymakers and technologists must collaborate to maximize benefits and mitigate risks [42,43].

6 Future Directions and Research Challenges

Despite significant progress, the current landscape of Agentic AI still exhibits substantial gaps in theory, practice, and governance. This section highlights these research gaps and proposes future directions for achieving scalable, trustworthy, and socially aligned Agentic AI systems. Future research into agentic AI must address several key areas:

a) *Generalization and Transfer Learning*: Building agents that can abstract knowledge and adapt behaviors across varied domains, minimizing the need for domain-specific engineering.

b) *Multi-Agent Coordination and Social Intelligence*: Developing mechanisms for effective negotiation, cooperation, and competition among heterogeneous agents and human collaborators.
c) *Explainable and Trustworthy Agentic AI*: Advancing interpretability and transparency, especially in safety-critical contexts, to foster trust and accountability.
d) *Human-AI Alignment*: Ensuring agentic AI systems reliably pursue objectives that reflect human values, intentions, and ethical constraints.
e) *Robustness and Resilience*: Enhancing agentic AI's ability to withstand adversarial inputs, system failures, and environmental uncertainty.
f) *Ethical Governance and Regulation*: Creating adaptive legal frameworks and industry standards to guide the responsible deployment of agentic AI, safeguarding public interest and societal well-being.
g) *Sustainable and Inclusive AI*: Ensuring the benefits of agentic AI are distributed equitably, and negative externalities are mitigated across societies.

Table 3 represents Agentic AI's technical, ethical, and policy challenges.

Table 3 Technical, ethical, and policy challenges

Category	Challenge	Description
Technical	Autonomy Control	Ensuring systems operate within intended bounds and can be reliably directed or halted by humans.
Technical	Goal Alignment	Aligning AI objectives with complex, evolving human values to avoid unintended consequences.
Technical	Robustness & Reliability	Building AI that performs safely and reliably in unpredictable real-world scenarios.
Ethical	Accountability	Determining who is responsible for AI actions, especially in cases of autonomous decision-making.
Ethical	Transparency	Making AI decision processes interpretable and explainable to users and stakeholders.
Ethical	Bias & Fairness	Ensuring AI does not perpetuate or exacerbate social biases or discrimination.
Policy	Regulatory Frameworks	Establishing laws and guidelines tailored to the risks and needs of agentic AI systems.
Policy	International Coordination	Coordinating standards and enforcement across jurisdictions to prevent regulatory gaps.
Policy	Liability & Insurance	Defining liability for damages or harm caused by autonomous AI agents.

7 Conclusion

Agentic AI stands at the forefront of technological innovation, poised to revolutionize core aspects of modern life. Its applications span healthcare, transportation, robotics, finance, education, and infrastructure, delivering both unprecedented opportunities and new challenges. As these systems mature, deliberate efforts are needed to ensure their development and deployment align with social values, human rights, and global sustainability. Continued interdisciplinary collaboration, vigilant oversight, and adaptive governance will be crucial in unlocking the full promise of agentic AI for society. Future research should focus on developing robust evaluation methods for agentic AI behavior and mechanisms for transparent and accountable decision-making. Exploring cross-disciplinary approaches will also be essential to anticipate emerging risks and maximize the societal benefits of these technologies.

Bibliography

[1] S. Goyal and G.K. Goyal, "Artificial Neural Network Simulated Elman Models for Predicting Shelf Life of Processed Cheese". Int. J. of App. Metaheur. Comp., Vol.3, No.3, pp.20-32, 2012. https://doi.org/10.4018/jamc.2012070102

[2] S. Goyal, "Artificial Neural Networks in Fruits: A Comprehensive Review". Int. J.of Image, Grap. and Sig. Process.,Vol.6, No.5, pp. 53, 2014. https://doi.org/10.5815/ijigsp.2014.05.07

[3] R. Kaur, D. Gabrijelčič, and T. Klobučar, "Artificial intelligence for cybersecurity: Literature review and future research directions", Info. Fusion, Vol.97, pp. 101804, 2023. doi: https://doi.org/10.1016/j.inffus.2023.101804

[4] S. Latif, F. Shamshad, H. S. Ali, F. Pervez, E. Cambria, and H. Cuayáhuitl, "A survey on deep reinforcement learning for audio-based applications", Artificial Intelligence Review, Vol. 56, No. 3, pp. 2193–2240, 2022. https://doi.org/10.1007/s10462-022-10224-2

[5] S. Mo, X. Pei, and C. Wu, "Safe reinforcement learning for autonomous vehicle using Monte Carlo tree search", IEEE Transactions on Intelligent Transportation Systems, Vol. 23, No. 7, pp. 6766–6773, 2022. https://doi.org/10.1109/tits.2021.3061627

[6] M. M. Karim, Q. Qu, D. H. Van, S. Khan, and Y. Kholodov, "AI agents meet blockchain: A survey on secure and scalable collaboration for multi-agents", Future Internet, Vol. 17, No. 2, p. 57, 2025. https://doi.org/10.3390/fi17020057

[7] N. Bougueroua, A. Derhab, S. Mazouzi, M. Belaoued, N. Seddari, and A. Bouras, "A survey on multi-agent based collaborative intrusion detection systems", Journal of Artificial Intelligence and Soft Computing Research, Vol. 11, No. 2, pp. 111–142, 2021. https://doi.org/10.2478/jaiscr-2021-0008

[8] P. Hitzler, A. Eberhart, M. Ebrahimi, M. K. Sarker, and L. Zhou, "Neuro-symbolic approaches in artificial intelligence", National Science Review, Vol. 9, No. 6, 2022. https://doi.org/10.1093/nsr/nwac035

[9] R. Sajja, M. Cikmaz, D. Cwiertny, Y. Sermet, and I. Demir, "Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education", Information, Vol. 15, No. 10, p. 596, 2024. https://doi.org/10.3390/info15100596

[10] T. Taniguchi, H. Yamakawa, T. Nagai, K. Doya, M. Sakagami, M. Suzuki, T. Nakamura, and A. Taniguchi, "A whole brain probabilistic generative model: Toward realizing cognitive architectures for developmental robots", Neural Networks, Vol. 150, pp. 293–312, 2022. https://doi.org/10.1016/j.neunet.2022.02.026

[11] P. Radanliev, "Artificial intelligence: Reflecting on the past and looking towards the next paradigm shift", Journal of Experimental & Theoretical Artificial Intelligence, Vol. 37, No. 7, pp. 1045–1062, 2024. https://doi.org/10.1080/0952813x.2024.2323042

[12] L. W. Y. Yang, X. Lei, X. Zhang, M. Yan, Y. Liu, D. S. W. Ting, L. L. Foo, and W. Y. Ng, "Deep learning-based natural language processing in ophthalmology: Applications, challenges and future directions", Current Opinion in Ophthalmology, Vol. 32, No. 5, pp. 397–405, 2021. https://doi.org/10.1097/icu.0000000000000789

[13] Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. Artificial Intelligence Review, 57(4). https://doi.org/10.1007/s10462-024-10721-6

[14] E. Dilek and M. Dener, "Computer vision applications in intelligent transportation systems: A survey", Sensors (Basel, Switzerland), Vol. 23, No. 6, p. 2938, 2023. https://doi.org/10.3390/s23062938

[15] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction", MIT Press, 2018. https://doi.org/10.1016/S0893-6080(99)000982

[16] J. Chen, "Generative AI for Social Engineering Attack Simulation," IEEE Trans. Inf. Forensics and Sec., Vol.16, pp. 4871-4883, 2021. https://doi.org/10.13140/RG.2.2.19118.32326

[17] N. Bougueroua, S. Mazouzi, S. Belaoued, N. Seddari, A. Derhab, and A. Bouras, "A survey on multi-agent based collaborative intrusion detection systems", J. of A. I. and Soft Comp. Res., Vol.11, No.2, pp. 111-142, 2021 https://doi.org/10.2478/jaiscr-2021-0008

[18] A. O. Hashesh, M. M. Fouda, R. M. Zaki, S. Hashima, K. Hatano, and A. S. T. Eldien, "AI-enabled UAV communications: Challenges and future directions", IEEE Access, Vol. 10, pp. 92048–92066, 2022. https://doi.org/10.1109/access.2022.3202956

[19] L. Hughes, Y. K. Dwivedi, T. Malik, M. Shawosh, M. A. Albashrawi, I. Jeon, V. Dutot, M. Appanderanda, T. Crick, R. De', M. Fenwick, S. M. Gunaratnege, P. Jurcys, A. K. Kar, N. Kshetri, K. Li, S. Mutasa, S. Samothrakis, M. Wade, and P. Walton, "AI agents and agentic systems: A multi-expert analysis", Journal of Computer Information Systems, Vol. 65, No. 4, pp. 489–517, 2025. https://doi.org/10.1080/08874417.2025.2483832

[20] D. Gupta, "Generative AI and Deep fakes: Ethical Implications and Detection Techniques", J. of Sci., Tech. and Engg. Res., Vol.1, No.1, pp. 45-56, 2024. https://doi.org/10.64206/21rgkc40

[21] T. Ha, T.K. Dang, H. Le and T.A. Truong, "Security And Privacy Issues in Deep Learning: A Brief Review. SN Computer Science", Vol.1, No.5, pp. 253, 2020. https://doi.org/10.1007/s42979-020-00254-4

[22] S. H. Thorat, "Agentic AI for Customer Service and Contact Center Solutions". J. of Com. Science and Tech. Stud.", Vol.7, No.8, pp. 444-451, 2025. https://doi.org/10.32996/jcsts.2025.7.8.49

[23] Y. Zhang and K. Siau, "Meta-entrepreneurship: An Analysis Theory On Integrating Generative AI, agentic AI, and metaverse for entrepreneurship". J. of G. Info. Mgmt., Vol. 32, No. 1, pp.1-21, 2024. https://doi.org/10.4018/JGIM.364094

[24] U.M. Borghoff, P. Bottoni and R. Pareschi, "Human-Artificial Interaction In The Age of Agentic AI: a system-theoretical approach". Fron. in Hum. Dyn., Vol. 7, pp. 1579166, 2025. https://doi.org/10.3389/fhumd.2025.1579166

[25] S. Sivakumar, "Agentic AI in predictive AIOPs: Enhancing It Autonomy and Performance. Int. J. of Sci. Res. and Mgmt., Vol. 12, No. 11, pp. 1631-1638, 2024 https://doi.org/10.18535/ijsrm/v12i11.ec01

[26] C. Chawla, S. Chatterjee, S.S. Gadadinni, P. Verma, and S. Banerjee, "Agentic AI: The building blocks of sophisticated AI business applications," J. of AI, Rob. & Workp. Auto, Vol. 3, No.3, pp.1-15, 2024. https://doi.org/10.69554/XEHZ1946

[27] I. D. S Portugal, P. Alencar and D. Cowan, D., "An Agentic AI-Based Multi-Agent Framework For Recommender Systems. In IEEE Int. Conf. on Big Data, pp. 5375-5382, 2024. IEEE. https://doi.org/10.1109/BigData62323.2024.10825765

[28] V. Shankar, "Managing The Twin Faces of AI: A Commentary on "Is AI Changing The World For Better or Worse?". J. of Macromark., Vol. 44, No.4, pp. 892-899, 2024. https://doi.org/10.1177/02761467241286483

[29] A.B. Arrieta, N. Díaz-Rodríguez, J.D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities And Challenges Toward Responsible AI", Info. Fusion, Vol. 58, pp. 82-115, 2020. https://doi.org/10.1016/j.inffus.2019.12.012

[30] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair and Y. Bengio, "Generative adversarial nets", Adv. in Neural Info. Process. Sys., Vol. 27, pp. 1-9, 2014. https://doi.org/10.1145/3422622

[31] V. Garg, "Designing the Mind: How Agentic Frameworks Are Shaping the Future of AI Behavior", J. of Comp. Sci. and Tech. Stud., Vol.7, No.5, pp. 182-193, 2023. https://doi.org/10.32996/jcsts.2025.7.5.24

[32] Y. Tu, J. Chen and C. Huang, "Empowering personalized learning with generative artificial intelligence: Mechanisms, challenges and pathways", Fron. of Digital Edu., Vol.2, No. 2, pp. 1-18, 2025. https://doi.org/10.1007/s44366-025-0056-9

[33] T. Miller, "Explanation In Artificial Intelligence: Insights From The Social Sciences", Art. Intel., Vol. 267, pp. 1-38, 2019. https://doi.org/10.1016/j.artint.2018.07.007

[34] H. F. Rashvand, K. Salah, J.M.A. Calero and L. Harn, "Distributed Security For Multi-Agent Systems–Review And Applications", IET Info. Sec., Vol. 4, No.4, pp. 188-201, 2010. https://doi.org/10.1049/iet-ifs.2010.0041

[35] C. Yinka-Banjo and O..A. Ugot, "A Review of Generative Adversarial Networks and Its Application In Cybersecurity. Art. Intel. R.", Vol. 53, No. 3, pp. 1721-1736, 2020. https://doi.org/10.1007/s10462-019-09717-4

[36] D. Cohen, D. Te'eni, I. Yahav, A. Zagalsky, D. Schwartz, G. Silverman and J. Makowski, "Human–AI Enhancement of Cyber Threat Intelligence", Int. J. of Info. Sec., Vol. 24, No. 2, pp. 99, 2025. https://doi.org/10.1007/s10207-025-01004-4

[37] D. Cohen, D. Te'eni, I. Yahav, A. Zagalsky, D. Schwartz, G. Silverman and J. Makowski, "Human–AI Enhancement of Cyber Threat Intelligence", Int. J. of Info. Sec., Vol. 24, No. 2, pp. 99, 2025. https://doi.org/10.1007/s10207-025-01004-4

[38] L. Hughes, Y.K. Dwivedi, T. Malik, M. Shawosh, M.A. Albashrawi, I. Jeon, V. Dutot, M. Appanderanda, T. Crick, R. De', M. Fenwick, "AI agents and agentic systems: A multi-expert analysis", J. of Comp. Info. Sys., Vol. 26, pp. 1-29, 2025. https://doi.org/10.1080/1097198X.2025.2524286

[39] B. Zyoud and S.L. Lutfi, "Adapting Zero Trust: Information Security Cultural Factors Considerations in the UAE Context", Asia-Pac. J. of Info. Tech. & Mul., Vol. 13, No. 2, 2024. https://doi.org/10.17576/apjitm-2024-1302-09

[40] I. H. Sarker, M.H. Furhad and R. Nowrozy, "AI-driven cybersecurity: an overview, security intelligence modeling and research directions", SN Comp. Sci., Vol. 2, No. 3, pp.173, 2021. https://doi.org/10.1007/s42979-021-00557-0

[41] N. Papernot, "Practical Black-Box Attacks against Machine Learning," In Proc. ACM Asia Conf. on Comp. and Comm. Sec., 2017. https://doi.org/10.1145/3052973.30530

[42] L. Wang, "Autonomous Security Operations Centers: A Case Study," In Proc. IEEE Conf. on Comm. and Net. Sec., pp. 550-561, 2022. https://doi.org/10.63282/3050-922X.IJERET-V6I2P108

[43] V. Kumar and D.Sinha, "Synthetic attack data generation model applying generative adversarial network for intrusion Detection", Comp. & Sec, Vol. 125, pp. 103054, 2023. https://doi.org/10.1016/j.cose.2022.103054

Authors' Profile

Dr. Sumit Goyal has a PhD in Computer Science (AI). In the past, he has worked with top brands and clients, including IBM, Lufthansa, Tieto EVRY, Google, BMW, Volkswagen, ABN-AMRO, Danone, Apple, Henkel, Michelin, the National Dairy Research Institute, the Centre for Development of Advanced Computing, and others. He has published over 100 research and review papers in the areas of AI, Machine Learning, and Cloud Computing, which more than 2,000 researchers have cited. His areas of scientific interest include Machine Learning, Agentic AI, Generative AI, Cloud Computing, and Cyber Security.