International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 10 No. 4, 2024
https://theaspd.com/index.php

Agentic AI Systems: A Review of Architectures, Autonomy, and Ethical Implications

Naeem Sayyad[1]*, Hetavi Dave[2], Mayank Kathane[3], Rashmi Patel[4]
[1*]Department of Data Science (M.Tech), Mukesh Patel School of Technology Management and Engineering, NMIMS Deemed to be University - Vile Parle(W), Mumbai 400056
naeem.sayyad30@nmims.in
[2]Department of Data Science (M.Tech), Mukesh Patel School of Technology Management and Engineering, NMIMS Deemed to be University - Vile Parle(W), Mumbai 400056
hetavi.dave22@nmims.in
[3]Department of Data Science (M.Tech), Mukesh Patel School of Technology Management and Engineering, NMIMS Deemed to be University - Vile Parle(W), Mumbai 400056
mayank.kathane21@nmims.in
[4]Department of Data Science (M.Tech), Mukesh Patel School of Technology Management and Engineering, NMIMS Deemed to be University - Vile Parle(W), Mumbai 400056
Rashmi.patel@nmims.edu

Abstract
Artificial Intelligence (AI) is a new revolution in the development of intelligent systems, which enables the machine to perceive, reason, and take intelligent actions without much human interference. This review attempts to explore architecture principles, autonomy models, ethical concerns and application of agentic AI in terms of its potential to deliver environmental sustainability. It highlights the transition towards architectures based on advanced cognitive, modular, reinforcement learning-based, and Hybrid systems and points out the fact that these systems are able to be more flexible, scalable and make better decisions. The review explains how agentic AI has been used in climatic modelling, biodiversity monitoring, intelligent agriculture, sustainable cities, and disaster response and demonstrates how the technology has the capability of unleashing data driven ecological decision-making. Moreover, it takes care of governance, like accountability, transparency, and value alignment issues, peripheral to the prudent application of such technologies to ecologically fragile zones. Finally, the article even provides some critical paths to go in the future such as scalability of computations, reliability, explicability, and even ethical questions in critical contexts, but rather proposes future tracks in human ai cooperation, policy relations, and even sustainable invention. The agentic AI is a revolutionary force that could help in reducing the threat of the climatic predicament, resource management as well as environmental conservation within the global ecosystem through the incorporation of the technological advancement with environmental stewardship. This review points at the explosion of agentic AI in confronting environmental decision-making, climate sustainability, and ecological resilience.

Keywords: *Agentic AI systems, environmental sustainability, autonomous decision-making, climate intelligence, ethical governance*

1. INTRODUCTION

Artificial intelligence in agent form (AI) is a paradigm shift towards the construction of intelligent entities that autonomously plan, act and evolve to achieve complex tasks, without repeated human direction [1]. Unlike traditional AI architecture that is conditioned with fixed commands, agentic architecture provides the ability to dynamically learn, proactive thinking, and self-directed problem solving, which can adjust to various and unpredictable settings. These systems have found additional use in industrial applications, such as robotics and healthcare, autonomous systems, and business automation, as autonomy and flexibility are primary drivers of efficiency and effectiveness. Agents AI have a potential to deal with vast amounts of data, learn through interaction and enhance decision pipelines making it a key enabler of future-generation smart systems. Moreover, the fact that being reduced to more and more dependent on such systems opens serious ethical and social issues, namely the considerations of compatibility with human values, transparency and fairness of decision-making. The fact that agentic AI is taking the leading role in shaping the technological ecosystems in the industry and academia makes it central to examine the architectures and autonomy of agentic AI as well as the governance implications thereof.

Nevertheless, due to the impressive advancements in agentic AI, work and a design are fragmented, and unsatisfactory, in overcoming the problems of autonomous systems [2]. Conventional AI models are adept in synthetic, simple-task settings but fail when forced into the unknown, multi step and dynamic

real world. Solving the question of completely agentic systems is the problem of combating such issues as a high level of goal-setting, adaptation-based reasoning, ethical decision-making, and transitions that occur without glitches between a collection of self-governing agents and human stakeholders [3]. Also, studies have been done looking at both cognitive architecture, as well as reinforcement learning, but the technical frameworks have not been combined with corresponding ethical factors to allow a sustainable deployment [4]. Architectures, autonomy and ethical considerations are being conceived in current literature without any connection or interrelation and this creates gaps which impose limitations to understanding holistically and application to the real world [5]. So-called gaps in knowledge highlight the necessity to conduct a comprehensive review that would not only bring together existing knowledge but also outline the openings of future research and potential ethics models to enable the responsible use of agentic AI systems.

This study will feature the comprehensive summation of the current state of agentic AI in an account, findings of architecture design and taxonomies of autonomy and ethical framework can be integrated in a cohesive vision. The study will also aim to identify important knowledge gaps in the existing literature and present action recommendations that can guide future action in terms of both technical breakthroughs and ethical control of agentic AI systems given the vision of its responsible and sustainable use.

2. The Grounds of Agentic AI

2.1 Definition of Agentic AI

There are two variants of artificial intelligence: Agentic Artificial Intelligence is computer systems that have self-regulating decisions, adaptive learning, and goal-oriented behaviour without frequent interaction with humans [6]. The traditional AI that is programmed with pre-set directions differs with the agentic AI that is not programmed and exhibits intentionality in making proactive decisions on how to find the most effective procedures in the changing environments. These systems use advanced cognitive designs and reinforcement learning to arrive at autonomic behaviors whilst remaining present to their situation [7]. According to new frameworks, the purpose of agency is a key feature that is given priority the ability of the system to think, perceive, and act on its own and in accordance with the targeted objectives and social rules. Thus, agentic AI is a groundbreaking paradigm in the development of intelligent, flexible and responsible computational agents.

2.2 Features and Abilities of Agentic Systems

Among the capabilities that are unique to the AAIs are the autonomy, proactiveness, adaptability, and continual learning [8]. The ability to respond autonomously in uncertain situations rather than responding by pre-programmed tasks should be possible in them with the input of dynamic contextual information. The systems have mechanisms of self-improvement such that they can adjust the strategies they follow over time and maximise their performance by means of feedback actions. The agentic architectures also consider goal prioritisation, plans and strategic thinking and multi-step reasoning and so may be applicable to applications like robotics or finance and personalised medicine [9]. Because they can be separately responsibly operational, they are the critical ingredient in an intelligent infrastructure of the future.

2.3 Contrast with Conventional AI Architectures

Traditional AI solutions are restricted to rule-based programming and learning and task-specific models and, therefore limit their adaptability under real world conditions of decision-making. Adaptive reasoning, autonomous planning and the ability to learn throughout the life provides response to increased functional autonomy of an agentic AI, as data in Figure 1 reveals. Where conventional designs require a large load of human interaction, agent-based designs continually perceive the environment, update knowledge representations and adjust actions during run-time [10]. Also unlike other AI frameworks, agentic systems make decision-making ethically aligned and focused on usability that centers on the ethical norms of value-sensing design, which is possible to carry responsibly. This paradigm switch causes agentic AI to be able to handle the multi-step, multi-risk taskings well beyond the capacity of a tradition machine learning model.

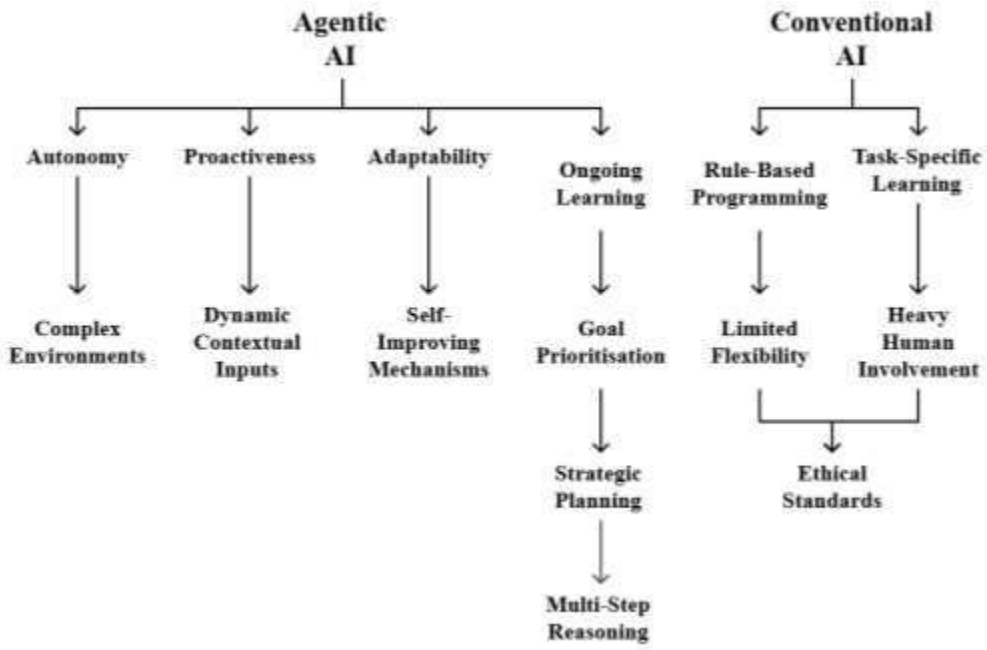

Figure 1: Agentic AI vs Conventional AI

3. Architectures of Agentic AI Systems

3.1 Cognitive and Modular Architectures

The agentic AI is based on cognitive and modular architectures where processing is organised in autonomous, logic units of processing [11]. Those architectures localise key cognitive capabilities, including planning, learning, and decision-making into specialised modules interacting with one another to increase their scalability and adaptability. With the integration of sub- symbolic learning and symbolic reasoning cognitive frameworks, these extend problem solving processes in changing worlds [12]. Besides, modularity permits flexibility in upgrades and task-specific customisation that can be applied in different fields like robotics and autonomous decision-support systems. Such an architecture can make agentic systems more autonomous and at human level intelligence.

3.2 Multi-Agent and Distributed Architectures

Multiagent and distributed systems allow widespread uses of agentic AI systems in terms of the networks of collaborating agents, where multiple independent agents interact in order to fulfil shared goals. These architectures use the communication syntaxes, coordination algorithms, and negotiation procedure that can assist in governing intricate choices in collaborative settings [13]. Distributed models allow independence of individual agents with the capacity to perform a collective process output. They also apply best where knowledge needs to be shared in real-time such as in smart grids, traffic optimisation tasks and financial predictions that involve high demand areas. With the composition of autonomous functionalities and the combination of co-operation platforms, the multi-agent structures enable competent sharing of resources and learning adaptation.

3.3 Reinforcement Learning-Based Architectures

The body of RL-based systems constitutes another major pillar of agentic AI as the architectures can be trained to find optimal action structures what is achieved by trial-and-error in the evolving environments. These designs are reward-based feedback signals to enhance decisions as a system in the future which will help to perform better and be adaptable. The use of advanced RL technologies, like deep learning and policy-gradient algorithms, means that agentic systems can be able to work on highly complex, multi-tasking problems without much human information. The applications run a continuum between autonomous robotics systems at one end to executing an adaptive recommendation at the other end; making RL-powered solutions highly extensible [14]. Induction of RL into agentic designs allows the systems to exhibit proactive reasoning and continuous amelioration without affecting their performance.

3.4 Hybrid and Hierarchical Models

Symbolic reasoning, neural learning, and the layered decisions found in hybrid and hierarchical architectures are integrated to enhance the flexibility and also the intelligence of agentic AI systems. The hybrid models combine both the knowledge-driven models as well as data-driven models that enable the systems to handle the structured reasoned inferences as well as flexibilities of context as indicated in Table 1. The structure decision processes of hierarchical architecture makes use of multiple layers, beginning with low-level perception to high-level planning, optimises a real-time task execution. These models find more and more applications in autonomous cars, healthcare diagnostics, and factory automation where reliability and precision are the most significant [15]. The combination of different paradigms, hybrid-hierarchical systems offer scalable, explainable and productive agentic AI systems.

Table 1. Comparison of Architectures in Agentic AI Systems

Architecture Type	Core Principle	Key Components	Capabilities	Applications	Advantages	References
Cognitive & Modular Architectures	Divide complex processes into specialised modules for reasoning, planning, and decision-making.	Knowledge representation, symbolic reasoning, neural learning, planning units	Enable perception, reasoning, and adaptive goal-driven behaviour	Robotics, autonomous decision-support, industrial automation	High scalability, customizability, and flexible module upgrades	[15]
Multi-Agent & Distributed Architectures	Use collaborative networks of autonomous agents that communicate and coordinate to achieve shared goals.	Agent communication protocols, coordination mechanisms, negotiation frameworks	Real-time multi-agent decision-making, distributed control	Smart grids, traffic optimisation, and financial forecasting	Scalability, parallel processing, and cooperative resource sharing	[13]
Reinforcement Learning-Based Architectures	Learn optimal strategies via trial-and-error and feedback-driven adaptation.	Reward-based feedback loops, deep RL, policy-gradient algorithms	Adaptive learning, autonomous planning, proactive decision-making	Robotics, personalised recommendations, adaptive control systems	Continuous improvement, adaptability, and minimal human	[11]

					interv entio n	
Hybrid Architectur es	Integrate symbolic reasoning with data-driven learning for adaptive intelligence.	Knowledge-driven models, neural networks, and contextual inference modules	Balance structured reasoning with contextual flexibility	Healthcare, enterprise AI, intelligent assistants	Com bines the accur acy of symb olic AI with the learni ng efficie ncy of neura l mode ls	[13]
Hierarchica l Architectur es	Organise decision-making processes into layers for optimised control and adaptability.	Low-level perception layers, mid-level planning, and high-level reasoning modules	Efficiently manage complex multi-stage tasks and dynamic goals	Autonomous vehicles, medical diagnostics, and industrial robotics	High precis ion, inter preta bility, and faster task execu tion	[14]
Integrated Agentic Systems	Combine multiple architectures to maximise autonomy and adaptability	Cognitive modules, multi-agent control, RL policies, hybrid reasoning layers	Enable self-evolving and goal-driven autonomy across environme nts	Multi-domain AI, AI governance frameworks, next-gen enterprise systems	High est adapt abilit y, real-time learni ng, and resou rce opti misat ion	[11]
Summary	Highlights the diversity of architectures enabling agentic AI	Varies across symbolic, hybrid, modular, and RL-driven paradigms	All models aim to enhance autonomy, adaptabilit y, and decision quality	Applications span robotics, healthcare, finance, industry, and enterprise		[13]

4. Autonomy Levels in Agentic Systems

4.1 Autonomy Taxonomy

The taxonomy of autonomy in agentic AI defines the gradual increase of independence that can be presented in a smart system, ranging between the automatized rule-based work to intelligence-based decision-making [16]. Systems in lower levels have followed a pre-determined set of instructions with high levels of human supervision, whereas those in higher levels involve contextual learning and sensitive thinking. Sophisticated agentic models integrate prioritisation of goals, assessment of situations and errors that correct themselves to operate in dynamic environments. It is necessary to have this taxonomy to test the capability of the system, transparency and establishment of requirements within the regulating organs. By defining these levels of autonomy we can build architectures that can balance out machine intelligence and with human intervention.

4.2 Decision-Making and Goal-Setting Mechanisms

Agentic AI requires decision-making based on their capability to establish a goal, evaluate a decision, and put plans into operation with the state of the surroundings at a time. The modern agentic systems depend on reinforcement learning, probabilistic reasoning and predictive analytics in order to achieve good performance across diverse settings [17]. Such systems can accomplish goal-setting functions, such as prioritization, optimal resource management and adaption to variably changing constraints without the need of a person to oversee them. Ethical values, such as fairness and responsibility, are now becoming a part of these frameworks to ensure that a value-driven decision-making process takes place. As such, goal-directed reasoning architectures would allow agentic systems to be flexible, reliable and evolve with dynamic intelligence in the high-stakes environment [18].

4.3 Human-in-the-Loop vs. Fully Autonomous Systems

Agentic AI systems exist on a spectrum of control, spanning human-in-the-loop (HITL) systems, where humans directly oversee, to fully autonomous models that can self-govern and execute [19]. HITL architectures increase reliability and ethical responsibility through the integration of human judgment in important decision-making pipelines, as shown in Figure 2. Contrarily, completely autonomous systems use sophisticated learning architectures to autonomously plan, perform, and improve upon tasks within dynamic scenes [20]. But eliminating human supervision identifies responsibility, safety, and governance issues. The balance between autonomy and human intervention must be appropriately struck in order to reduce harms and support trustworthy AI deployment.

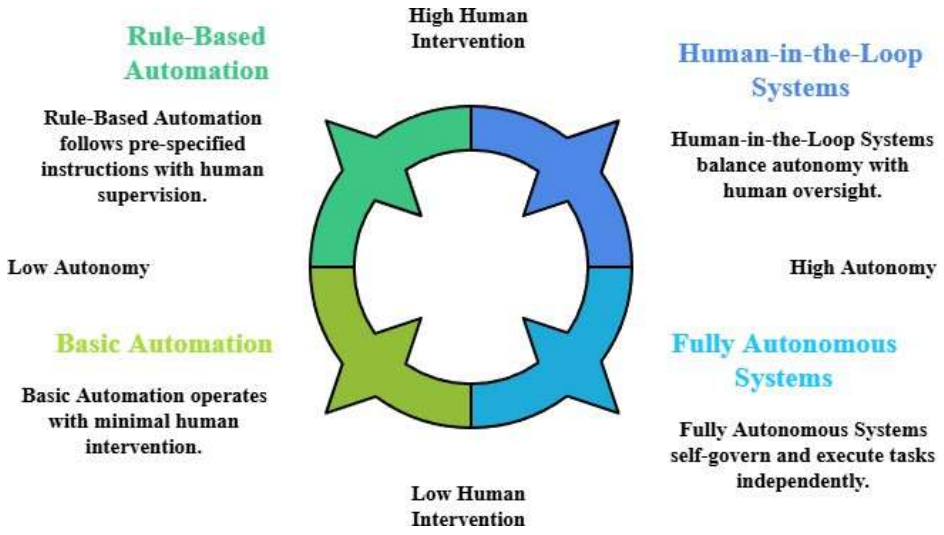

Figure 2: Autonomy Levels in Agentic Systems

5. Applications of Agentic AI

5.1 Robotics and Autonomous Vehicles

Agentic AI has a revolutionary impact on robotics and autonomous vehicles, allowing systems to achieve real-time perception, adaptive decision-making, and coordinated control in dynamic environments [21].

With the incorporation of sophisticated reinforcement learning and sensor-fusion methods, agentic robots are able to move through complex landscapes and manage uncertain situations with very little human intervention. In autonomous transport, agentic architectures govern path planning, obstacles, and traffic forecasting, enhancing safety and efficiency [22]. The systems also encourage shared decision-making by two or more agents, and vehicle-to-vehicle communications to make the best choices of routes. Antecedent uses that of antecedent AI are, then, automation, resilience, and operational smartness of modern mobility systems.

5.2 Healthcare and Personalised Medicine

A field like medicine that requires personalised medicine, diagnostics, and clinical decision-making is being transformed because of agentic AI, which enables patient care delivery. Such systems, utilizing predictive modelling, multimodal data integration, and adaptive learning may provide their users with customised recommendations of treatment based on the unique profile of their patients [23]. Autonomous surgical robots, live monitoring of diseases and intelligent drug management can also be performed in an agentic framework to make clinical outcomes more effective. Ethical imperatives, including data privacy, informed consent, and transparency, remain the center-of-interest to their deployment, assuring the patient safety and trust. The ability of the agentic healthcare systems to revise medication approaches and drive towards precision medicines on an ongoing basis is also supported.

5.3 Finance, Security, and Enterprise Systems

Rapidly evolving financial markets, cybersecurity, and enterprise operations are being reshaped by Agents AI in such a way that they attract adaptive automation, fraud detection, and instant decision-making. In the financial services agentic models examine financial trends, default risks, and portfolio optimisation methodologies to enhance efficiency and accuracy of the predictions [24]. The speed at which cybersecurity measure are taken by agentic agents to spot anomalies, thwart threats and mitigate breaches is faster than traditional defence mechanisms. In business ecosystems, the genius systems to achieve simplified operations are the agentic architectures; the intelligent allocation of resource in combination with dynamic workflow management. The features enable organisations to ensure resilience and competitiveness in a way of addressing the sophistication of modern digital infrastructures across the sectors.

5.4 Creative and Knowledge-Intensive Domains

Possible applications Knowledge-intensive and creative industries: one of the applications seen in knowledge-intensive and creative industries is provision of innovations in research, content creation and design. The technology is capable of generating works of art, music, scientific models and designs of architectural structures independently through symbolic thinking intertwined to generative components [25]. In higher education and research and development, agentic platforms enable acceleration of knowledge discovery, hypothesis testing and literature synthesis, enabling rapid search over large spaces of complexity, as evident in Table 2. In leisure and entertainment, the AI makes authentic interactive narratives or recommendation possible and enables the creation of interactivity in digital spaces. These systems will allow industries to realise greater productivity and creative ability through cooperation of people and autonomy to deliver knowledge based workflows in a revolutionary way.

Table 2. Applications of Agentic AI Across Diverse Domains

Application Domain	Core Objective	Key Technologies Used	Capabilities Enabled	Practical Implementations	Advantages	References
Robotics & Autonomous Vehicles	Enable real-time perception, adaptive control, and coordinated navigation	Reinforcement learning, sensor fusion, and path-planning algorithms	Dynamic decision-making, vehicle-to-vehicle communication, and collaborative routing	Self-driving cars, autonomous drones, robotic delivery systems	Improves safety, efficiency, and scalability in complex mobility systems	[22]

Healthcare & Personalised Medicine	Deliver patient-specific diagnostics, treatment planning, and precision medicine.	Predictive modelling, multimodal data integration, and autonomous surgical robotics	Personalised care recommendations, disease monitoring, and drug optimisation	AI-powered diagnostics, surgical robots, and remote patient monitoring systems	Enhances treatment accuracy, reduces risks, and promotes precision healthcare	[24]
Finance & Cybersecurity	Automate complex decision-making for financial stability and secure digital ecosystems.	Deep learning, anomaly detection, predictive analytics, and autonomous risk assessment	Fraud detection, real-time portfolio optimisation, and autonomous threat response	Banking analytics, payment security, stock-trading bots	Enables fast responses, improved compliance, and robust threat mitigation	[25]
Enterprise Systems	Optimise resource allocation and manage workflows in dynamic environments	Multi-agent frameworks, intelligent process automation, and adaptive learning algorithms	Automates operational decision-making and task coordination across distributed systems	Intelligent ERP platforms, supply chain optimisation, and workforce management tools	Improves productivity, efficiency, and enterprise resilience	[22]
Creative Industries	Enhance knowledge-intensive innovation and autonomous content creation	Generative AI models, symbolic reasoning, LLM-driven design frameworks	Produces new art, music, scientific models, and architectural concepts autonomously	Interactive storytelling, AI-generated art, automated media production	Encourages creativity, personalisation, and user engagement	[23]
Research & Knowledge Discovery	Accelerate literature synthesis and scientific discovery through automation.	Semantic reasoning, hypothesis generation engines, and multi-agent knowledge integration	Identifies gaps, designs experiments, and optimises R&D processes	AI-assisted research platforms, drug discovery simulations	Speeds up knowledge generation and enhances decision accuracy	[25]
Summary	Highlights agentic AI's transformative impact across critical sectors	Combines RL, generative AI, multi-agent systems, and predictive analytics	Enhances decision-making, autonomy, and adaptability	Foundational in mobility, healthcare, finance, enterprise, and creative domains	Enables scalable innovation, ethical deployment, and human-AI collaboration	[21]

6. Environmental Applications and Sustainability Implications of Agentic AI

6.1 Climate Modelling and Environmental Monitoring

AgTech AI is transforming climate assessment due to the potential to perform real-time massive environmental data sets collected using satellites, IoT sensors and remote sensors. These systems are able to forecast extreme weather occurrence with reinforcement learning and multi-agent frameworks, and monitor ecosystem by use of reinforcement learning and multi-agent frameworks. The model of agentic

maximised expenditure of resources, implemented greenhouse gas surveillance, and adaptive climate risk mitigation. By autonomously computing complex environment ILs, the agentic AI can help scientists/policymakers make empirical and evidence-based environmental decisions, increasing the resilience and sustainability adaptability of the world to its environment at regional and planetary levels. As an example, such an agentic AI framework in Bangladesh considers satellite imagery, IoT sensor measurements, and reinforcement learning models to achieve high levels of accuracy in modeled flood patterns [26]. These revelations have been applied by local governments as an early warning to commence evacuation plan and disaster preparedness measures to alleviate possible environmental and human casualties.

6.2 Smart Agriculture and Resource Maximisation

Agentic AI leads the innovation of smart agriculture through the integration of precision farming technologies, soil health analysis, and sensor-driven irrigation systems to optimise crop yields while reducing resource loss [27]. These systems automatically adjust to climatic changes, maximise water allocation, and minimise reliance on toxic pesticides, encouraging ecologically friendly cultivation techniques. Agentic AI, through multi-agent coordination, examines real-time farming data, thus maximising land productivity and facilitating enhanced farmers' decision-making. Through maximisation of resource use, minimisation of ecological degradation, and facilitation of climate-resilient practices, agentic AI provides a scalable solution to food security in tandem with global sustainability and conservation. A recent pilot project in India implemented an agentic AI-powered smart irrigation system that monitors soil moisture levels and climatic variables in real time [28]. By autonomously adjusting water usage through adaptive learning, the system reduced water consumption by nearly 30% while increasing crop yields, demonstrating the role of AI in sustainable agriculture.

6.3 Urban Sustainability and Biodiversity Conservation

Autonomous drone technology using agentic AI and sensor-based monitoring facilitates year-round observation of wildlife, deforestation, and habitat destruction, thus supporting biodiversity conservation. The systems monitor, analyse, and report detailed ecological data to inform species conservation strategies and assess ecosystem well-being. Multi-agent systems in urban settings enhance sustainability by fine-tuning energy networks, traffic patterns, and waste management networks, resulting in lower carbon emissions and a cleaner city life. Through the synchronisation of intelligent automation and adaptive decision-making, agentic AI facilitates conservation operations, enhances sustainable city planning, and develops resilient urban ecosystems with the ability to fulfil long-term environmental and societal needs. In the Amazon rainforest, reinforcement learning-based agentic AI drones are deployed to track patterns of illegal deforestation and monitor biodiversity loss [29]. By autonomously processing large-scale satellite imagery and environmental data, these systems provide real-time alerts to conservation authorities, enabling faster interventions and better preservation of fragile ecosystems.

6.4 Disaster Response and Environmental Governance

Agentic AI greatly enhances disaster readiness through the examination of environmental data to predict floods, droughts, wildfires, and hurricanes with increased accuracy. These systems use multi-agent decision-making frameworks to maximise emergency response tactics and direct resource allocation in areas of high risk [30]. Agentic AI also helps to inform environmental governance through data-based policy recommendations that maintain technological effectiveness while ensuring ecological sustainability, as shown in Figure 3. Such intelligent systems make it possible to achieve the development of adaptive environmental policy through greater transparency and stakeholder involvement. The agentic AI can, therefore, serve as a foundation of diminishing natural mishaps, climate-resilience and sustainable utilization of global assets.

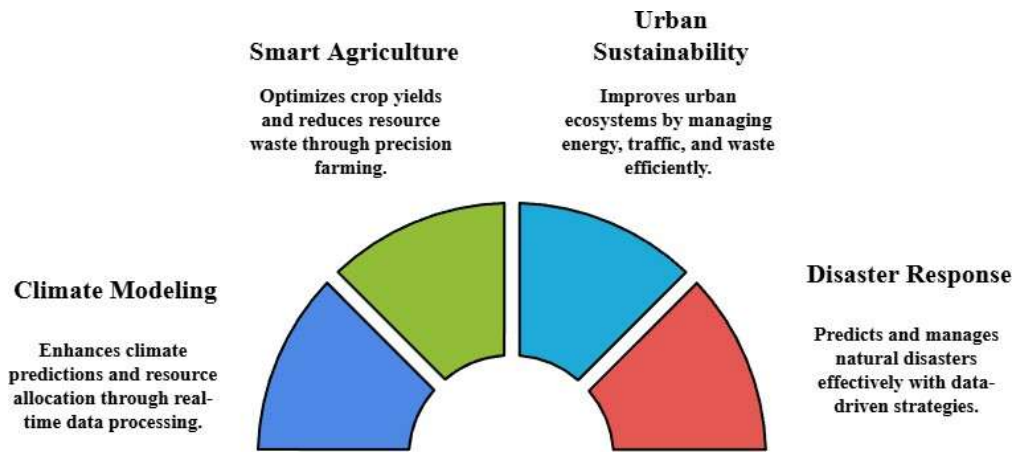

Figure 3: Agentic AI Environmental Impact

7. Ethical, Social, and Governance Implications

7.1 Accountability and Transparency

One critical aspect of agentic AI systems is accountability and transparency so that people can trust and rely on the responsible deployment of agentic AI systems. The mechanism of independent decision making carried out by such systems directly influences human life and it demands structured explanations of processes and mechanisms of reasoning [31]. Open reporting structures and XAI frameworks are also essential in order to monitor actions and results. Besides, the implantation of auditability controls aids in creation of joint accountability amongst developers, organisations, and regulators. Elevated trustworthiness Not only does greater openness complement system resiliency, but it also contributes to having harmonized AI-related goals with human values, which eases stakeholder trust and regulatory adherence in high-threat sectors.

7.2 Bias, Fairness, and Value Alignment

Algorithmic bias, fairness, and value alignment are the problems that AGI AI systems should address to deliver equal outcomes across populations. A discriminating pattern may appear through biases present in training data or algorithms used to make decisions and erode collective trust in the social fabric [32]. The consideration of the value-sensitive designing framework enables the developers to align system behaviours to ethics and the norms of society. Further, there is fairness since there is a continuous evaluation of the data, audit of the model, and stakeholders when developing and implementing it. Action on that, agentic AI can counter systemic injustices and build decision-making pipelines aiming at advancing human rights and diversity [33].

7.3 Safety, Control, and Human Supervision

To ensure safety, control, and guaranteed intensive human control of agentic AI systems, it becomes more necessary these systems will increasingly be more discreetly autonomous. Without any limits, autonomy may lead to unwanted actions that will inflict economic, environmental, or ethical losses. The value alignment technique, fail-safe modes, and Human-in-the-loop (HITL) architecture are used to trade between autonomy and operational safety [34]. Further, the application of agentic AI in requirements with high-security such as health and transportation ought to be under strict real-time surveillance to prevent failure and ensure accountability. By the introduction of open control structures, firms can make sure that operations are within the appropriate ethical and safety threshold.

7.4 Regulatory and Policy Considerations

The increase in the implementation of agentic AI brings a necessity to implement robust regulatory policies and guidelines to control autonomy, responsibility, and ethics. Existing regulatory systems are known to be slow in responding to a fast moving technological environment, necessitating flexible multi-stakeholder governance models [35]. International policy efforts now concentrate on data governance, explainability standards, and ethical AI certifications to promote responsible deployment across industries, as shown in Table 3. Regulatory functions should similarly mitigate difficulties such as cross-

border accountability, attribution of liability, and transparency obligations. By harmonising policy frameworks and technical innovation, governments and institutions can advance safe and fair application of agentic AI systems while limiting societal risks.

Table 3. Ethical, Social, and Governance Implications of Agentic AI Systems

Dimension	Core Issue	Underlying Cause	Impact on Agentic AI Systems	Proposed Solutions	Future Research Directions	References
Accountability & Transparency	Lack of clear explanations for autonomous decisions	Complex decision pipelines and opaque models	Reduced trust, limited adoption, and compliance risks	Implement Explainable AI (XAI) frameworks, open reporting, and audit mechanisms	Develop standardised accountability metrics and cross-industry transparency frameworks	[35]
Bias, Fairness & Value Alignment	Discriminatory patterns and unequal outcomes	Biased datasets and incomplete ethical integration	Compromised social equity, undermined credibility, and user dissatisfaction	Use value-sensitive design frameworks, diverse datasets, and continuous auditing	Establish global fairness benchmarks and inclusive AI development guidelines	[32]
Safety, Control & Human Oversight	Risk of unintended or harmful autonomous actions	Over-reliance on self-learning without proper safeguards	Economic losses, ethical violations, and system failures	Deploy fail-safe mechanisms, human-in-the-loop (HITL) models, and real-time monitoring	Research adaptive safety architectures that balance autonomy with oversight	[31]
Regulatory & Policy Considerations	Insufficient laws to manage rapidly evolving AI technologies	Slow legal adaptation and global policy fragmentation	Unclear accountability, liability disputes, and inconsistent ethical enforcement	Create multi-stakeholder governance frameworks and international AI compliance standards	Develop unified regulatory policies balancing innovation, ethics, and societal safety	[33]
Ethical Responsibility	Misalignment of AI objectives with human values	Focus on optimisation without societal context	Erosion of public trust and potential misuse in sensitive sectors	Define clear ethical boundaries and conduct regular impact assessments	Investigate frameworks for value-based AI goal alignment	[31]
Cross-Border Governance	Lack of global AI accountability and cooperation	Differing national regulations and compliance structures	Conflicts over data privacy, transparency, and liability	Establish international treaties and shared ethical standards for agentic AI	Research into global governance frameworks ensuring equity and inclusivity	[33]
Summary	Agentic AI introduces multi-dimensional	Driven by rapid AI advancement	Impacts safety, fairness, governance,	Requires interdisciplinary solutions combining	Focus on responsible, transparent, and inclusive	[31]

	ethical, legal, and social challenges.	and complex autonomy	and societal well-being	ethics, law, and AI design	AI innovation	

8. Challenges and Future Research Directions

8.1 Scalability and Computational Limitations

One of the main challenges in creating agentic AI systems is attaining scalability at the expense of performance or efficiency. With complex architectures, multi-agent coordination, and real-time decision-making, the computational requirements of such systems are much higher. High-performance computing infrastructure and distributed computing are critical to support large-scale data processing, dynamic simulation, and adaptive learning [36]. But energy efficiency, processing rates, and resource usage, while desired, continue to be a recurring bottleneck. Optimised architectures and adaptive algorithms will be the future study's focus to provide scalable agentic frameworks that can work in varied, high-demand environments and remain accurate and responsive.

8.2 Robustness and Reliability

Robustness and reliability of agentic AI systems have to be ensured for them to be deployed in real-world, high-stakes situations. There is the possibility of these systems being subjected to uncertainty, opposing inputs and divergent conditions of operations, and they might produce inconsistent or unsafe behaviour [37]. It is important to design fault-tolerant frameworks that will be able to respond to unexpected failures and resolve the resistant tampering [38]. The permanence of long-term and the availability of performance must be undertaken regularly through continuous validation, testing and monitoring. It must also explore self-healing and subsequent-generation resilience efforts to ensure agentic AI is safe, reliable, and consistent across all dynamic application environments.

8.3 Interpretability and Explainability

A current open challenge is whether it can be done to increase the interpretability and explainability of decision-making in agentic AI systems. As these models become more and more complex, it becomes more difficult to understand how they do their reasoning. Transparency is one of the greatest obstacles towards trust, accountability, and ethics compliance. Integrating explainable AI (XAI) architectures, causal thoughts and clear visualisation approaches is going to heighten stakeholder confidence and authority acceptability [39]. Also, the study area of high priority is to improve the performance optimisation by explaining aspects. It is essential to design systems that can be interpreted because it is essential to guarantee the community that unmanned choices would exhibit human values, social morals, and safety requirements.

8.4 Ethical Challenges in High-Risk Areas

Giving agentic AI such crucial uses in sectors such as healthcare, defence, finance and autonomous transportation presents complex ethical issues. These systems are faced with circumstances involving conflicting objectives, fairness constraints and safety trade-offs. The formulation of responsibility and accountability within autonomous actions is in desperate need of definition, particularly in areas where such an action has direct implications on human lives [40]. Further, ensuring system behaviour to meet societal values, cultural expectations, and regulatory constraints requires multidisciplinary strategies as shown in Figure 4. Future studies need to formulate strong ethical structures that combine transparency, inclusivity, and accountability in order to promote agentic AI systems to behave responsibly while reducing harm and increasing benefits to society.

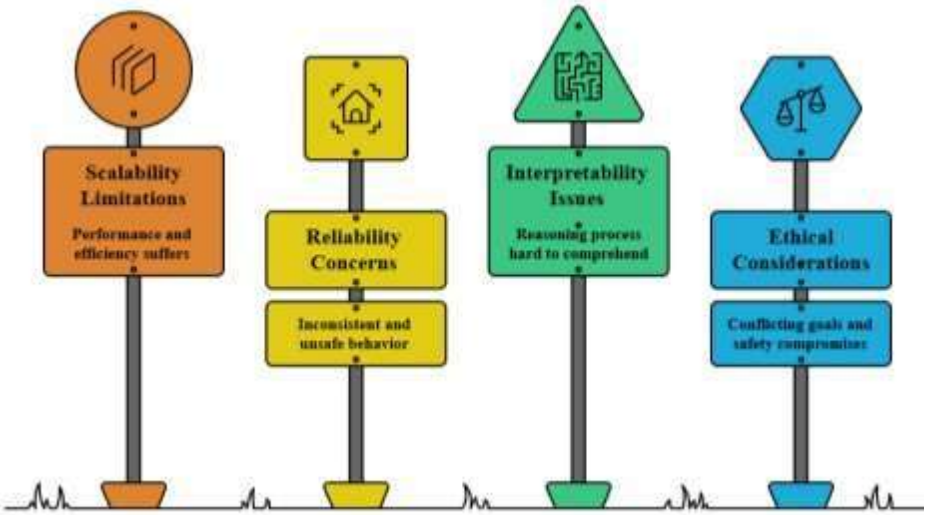

Figure 4: Agentic AI System Challenges

9. Future Directions

9.1 Synergy of Agentic AI and General Intelligence

Assembling agentic AI with artificial general intelligence (AGI) is a promising area in AI. Merger of existing agentic systems with AGI functionality would lead to generalised reasoning, learning adaptability, and decision-making across multiple contexts [41]. It would broaden system flexibility, making autonomous agents capable of handling a variety of tasks for various environments with continuous monitoring of ethical compliance. The problems persist in value embedding, safety limits, and governance models to achieve responsible deployment [42]. Scalable architecture and alignment methods are areas to which future studies need to concentrate, to bring agentic flexibility together with the larger cognitive abilities of AGI.

9.2 Models of Human-AI Collaboration

Progressive agentic AI requires creating cooperative frameworks where humans and smart systems are synergistic partners instead of substitutes. Human-AI partnership models seek to combine human judgment, contextual knowledge, and moral reasoning with agentic systems' computational abilities. Through the union of machine efficiency and human flexibility, such models can enhance decision-making in fields such as healthcare, finance, and autonomous robotics. Future studies need to investigate explainable interfaces and mutual learning processes to increase trust, transparency, and usability [43]. Such cooperative ecosystems will facilitate more responsible, robust, and human-centred deployments of agentic AI.

9.3 Emerging Paradigms and Research Opportunities

The accelerating development of agentic AI is compelling new paradigms in multi-agent systems, decentralised intelligence, and adaptive learning frameworks [44]. Future studies opportunities are in evolving self-optimising architectures that incorporate LLM-driven feedback loops and iterative refinement mechanisms for enhanced adaptability [45]. Furthermore, there is increased interest in integrating ethical thinking and value-sensitive designs into agentic pipelines to provide reliable deployment, as shown in Table 4. The interface of agentic AI, AGI, and web-based autonomous systems has the potential to redefine innovation in knowledge-intensive and operational spaces. Bridging these frontiers necessitates interdisciplinary research that integrates computing, ethics, and human-focused governance.

Table 4. Future Research Directions in Agentic AI Systems

Research Focus	Objective	Key Opportunities	Challenges	Proposed Solutions	Future Scope	References

Synergy of Agentic AI & AGI	Integrate agentic AI with artificial general intelligence for generalised cognition.	Unified reasoning, adaptability, and multi-context decision-making	Value embedding, safety constraints, and governance gaps	Develop scalable architectures, alignment strategies, and ethical compliance frameworks	Build hybrid systems capable of human-like intelligence with secure autonomous control	[41]
Human-AI Collaboration Models	Create frameworks where humans and agents act as cooperative partners	Combine human intuition, contextual knowledge, and AI's computational power	Lack of trust, interpretability, and explainability in collaboration models	Design mutual learning ecosystems, explainable interfaces, and ethical workflows	Enhance trustworthy, human-centred AI ecosystems across critical sectors	[43]
Emerging Paradigms in Multi-Agent Systems	Advanced distributed and decentralised intelligence for adaptive learning	Efficient coordination, resource sharing, and collective reasoning	System complexity, integration difficulties, and security risks	Use iterative refinement mechanisms and LLM-driven adaptive feedback loops	Enable self-optimising multi-agent ecosystems for dynamic environments	[45]
Integration of Ethical Design Frameworks	Embed value-sensitive designs into agentic pipelines for responsible AI	Ensure fairness, transparency, and inclusivity in decision-making	Managing bias, ethical inconsistencies, and conflicting objectives	Apply value alignment models and continuous ethical auditing	Foster sustainability and accountability in high-stakes decision environments	[43]
Web-Based Autonomous Ecosystems	Leverage agentic AI and AGI for next-generation internet and decentralised systems.	Autonomous content generation, intelligent information exchange, and collaborative platforms	Governance uncertainty, data security, and resource constraints	Create adaptive protocols for distributed control and knowledge sharing	Build intelligent, scalable, and resilient autonomous web ecosystems	[41]
Cross-Disciplinary Innovation	Merge AI, ethics, governance, and human sciences for holistic development	Enhance integration between cognitive architectures and societal values	Lack of unified frameworks and interdisciplinary coordination	Develop standardised methodologies combining computing, policy, and human-centred ethics	Drive innovation in sustainable AI governance and equitable deployment	[42]
Summary	Highlights the future pathways for evolving agentic AI systems	Emphasises intelligence integration, transparency, and adaptive architectures	Challenges lie in scalability, governance, and ethical alignment	Multi-stakeholder collaboration and iterative research are required	Enable responsible, scalable, and ethically compliant AI systems	[45]

10. CONCLUSION

This survey has examined the developing world of agentic AI systems, and they have their roots in structure, autonomy levels, ethical considerations, and heterogeneous real-world deployments. Compared to orthodox AI, an aspect giant shift is AGI AI because it has the ability of autonomous decision-making, adaptive reasoning, and self-directed goal management, among other complex environments. Using conversations about cognitive, modular, reinforcement learning-based and hybrid structures, the study demonstrates that advances in technology facilitate increased capacity and scalable systems. Moreover, the review also highlights important ethical issues of responsibility, transparency, eliminating bias and governance that remain very crucial in bringing forth the age of trust and social acceptability. Although they have made impressive progress, there are several issues that are not fully settled: scalability, robustness, interpretability, and ethical issues in high-stakes settings. Addressing them requires multidisciplinary research at the intersection of computing, cognitive science, ethics, and policy-making in order to make AI development responsible. Future work ought to be directed at merging agentic AI and general intelligence, the learning to collaborate models between humans and advanced AI, as well as applying harmonised regulatory frameworks on a global level. Agentic AI can empower ecological decision-making based on data and help put environmental resources at their most efficient point. This will help the world achieve sustainability, environmental protection and climate resiliency. Correlating technical breakthrough and moral accountability, agentic AI can revolutionise healthcare, robotics, finance, creative industries and enterprise ecosystems. However, its complete realisation depends on the development of open, safe, and value-driven systems that will balance autonomy with human control and allow sustainable innovation and maximum gains in society.

REFERENCES

1. Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous intelligence for complex goals–a comprehensive survey. IEEe Access.
2. Adamson, G., Havens, J. C., & Chatila, R. (2019). Designing a value-driven future for ethical autonomous and intelligent systems. Proceedings of the IEEE, 107(3), 518-525.
3. Beck, R., Dibbern, J., & Wiener, M. (2022). A multi-perspective framework for research on (sustainable) autonomous systems. Business & information systems engineering, 64(3), 265-273.
4. Cervantes, S., López, S., & Cervantes, J. A. (2020). Toward ethical cognitive architectures for the development of artificial moral agents. Cognitive systems research, 64, 117-125.
5. Chinta, P. C. R., & Karaka, L. M. (2020). Agentic AI and reinforcement learning: Towards more autonomous and adaptive AI systems. Journal for Educators, Teachers and Trainers https://jett. labosfor. com/index. php/jett/article/view/2699.
6. Dattathrani, S., & De', R. (2023). The concept of agency in the era of artificial intelligence: dimensions and degrees. Information Systems Frontiers, 25(1), 29-54.
7. Dodda, A. (2023). AI Governance and Security in Fintech: Ensuring Trust in Generative and Agentic AI Systems. American Advanced Journal for Emerging Disciplinaries (AAJED) ISSN: 3067-4190, 1(1).
8. Fang, J., Peng, Y., Zhang, X., Wang, Y., Yi, X., Zhang, G., ... & Meng, Z. (2025). A Comprehensive Survey of Self-Evolving AI Agents: A New Paradigm Bridging Foundation Models and Lifelong Agentic Systems. arXiv preprint arXiv:2508.07407.
9. Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. Machine learning and the city: Applications in architecture and urban design, 535-545.
10. Fritz, A., Brandt, W., Gimpel, H., & Bayer, S. (2020). Moral agency without responsibility? Analysis of three ethical models of human-computer interaction in times of artificial intelligence (AI). De Ethica, 6(1), 3-22.
11. Garg, V. (2025). Designing the Mind: How Agentic Frameworks Are Shaping the Future of AI Behavior. Journal of Computer Science and Technology Studies, 7(5), 182-193.
12. Hanif, M. A., Aleem, F. M., Anwar, F., Siddique, M., Iqbal, K., Sajjad, M., & Ahmad, G. (2025). BRINGING AUTONOMY AND COOPERATION TOGETHER: A COMPARISON OF AGENTIC AI SYSTEMS AND AI AGENTS. Spectrum of Engineering Sciences, 3(8), 59-68.
13. Hosseini, S., & Seilani, H. (2025). The role of agentic ai in shaping a smart future: A systematic review. Array, 100399.
14. Huggett, J. (2021). Algorithmic agency and autonomy in archaeological practice. Open Archaeology, 7(1), 417-434.
15. Jaggavarapu, M. K. R. (2025). The Evolution of Agentic AI: Architecture and Workflows for Autonomous Systems. Journal Of Multidisciplinary, 5(7), 418-427.
16. Jedlič'ková, A. (2025). Ethical approaches in designing autonomous and intelligent systems: a comprehensive survey towards responsible development. AI & SOCIETY, 40(4), 2703-2716.
17. Joshi, S. (2025). Advancing innovation in financial stability: A comprehensive review of ai agent frameworks, challenges and applications. World Journal of Advanced Engineering Technology and Sciences, 14(2), 117-126.
18. Joshi, S. (2025). Review of Autonomous and Collaborative Agentic AI and Multi-Agent Systems for Enterprise Applications.
19. Joshi, S. (2025). Review of autonomous systems and collaborative AI agent frameworks.
20. Karunanayake, N. (2025). Next-generation agentic AI for transforming healthcare. Informatics and Health, 2(2), 73-83.
21. Koubaa, A. (2025). From Pre-Trained Language Models to Agentic AI: Evolution and Architectures for Autonomous Intelligence.
22. Laitinen, A., & Sahlgren, O. (2021). AI systems and respect for human autonomy. Frontiers in artificial intelligence, 4, 705164.

23. Mansfield, D., & Montazeri, A. (2024). A survey on autonomous environmental monitoring approaches: towards unifying active sensing and reinforcement learning. Frontiers in Robotics and AI, 11, 1336612.
24. Methnani, L., Chiou, M., Dignum, V., & Theodorou, A. (2024). Who's in charge here? a survey on trustworthy ai in variable autonomy robotic systems. ACM computing surveys, 56(7), 1-32.
25. Mhlambi, S., & Tiribelli, S. (2023). Decolonizing AI ethics: Relational autonomy as a means to counter AI harms. Topoi, 42(3), 867-880.
26. Michael, K., Abbas, R., Roussos, G., Scornavacca, E., & Fosso-Wamba, S. (2020). Ethics in AI and autonomous system applications design. IEEE Transactions on Technology and Society, 1(3), 114-127.
27. Mojgani, R., Waelchli, D., Guan, Y., Koumoutsakos, P., & Hassanzadeh, P. (2023). Extreme event prediction with multi-agent reinforcement learning-based parametrisation of atmospheric and oceanic turbulence. arXiv preprint arXiv:2312.00907.
28. Nath, P., Moss, H., Shuckburgh, E., & Webb, M. (2024). RAIN: Reinforcement Algorithms for Improving Numerical Weather and Climate Models. arXiv preprint arXiv:2408.16118.
29. Pamisetty, A. (2024). Application of agentic artificial intelligence in autonomous decision making across food supply chains. Available at SSRN 5231360.
30. Patel, D. (2025). Agentic AI for Autonomous Decision-Making: Toward Ethical and Aligned Autonomy.
31. Pati, A. K. (2025). Agentic AI: A Comprehensive Survey of Technologies, Applications, and Societal Implications. IEEE Access.
32. Powell, D. (2019). Autonomous systems as legal agents: directly by the recognition of personhood or indirectly by the alchemy of algorithmic entities. Duke L. & Tech. Rev., 18, 306.
33. Priyadarshi, M. (2025). Autonomous AI Agents Transforming Enterprise Operations: From Static Automation to Intelligent Decision-Making Systems. Journal Of Multidisciplinary, 5(7), 863-870.
34. Prunkl, C. (2024). Human autonomy at risk? An analysis of the challenges from AI. Minds and Machines, 34(3), 26.
35. Rudd-Jones, J., Musolesi, M., & Pérez-Ortiz, M. (2025). Multi-Agent Reinforcement Learning Simulation for Environmental Policy Synthesis. arXiv preprint arXiv:2504.12777.
36. Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2025). Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges. arXiv preprint arXiv:2505.10468.
37. Shavit, Y., Agarwal, S., Brundage, M., Adler, S., O'Keefe, C., Campbell, R., ... & Robinson, D. G. (2023). Practices for governing agentic AI systems. Research Paper, OpenAI.
38. Shukla, A. K. (2025). From AI Agents to Agentic Intelligence: A Comparative Study of Autonomy, Adaptation, and Ethical Design. Adaptation, and Ethical Design (May 20, 2025).
39. Sonko, S., Adewusi, A. O., Obi, O. C., Onwusinkwue, S., & Atadoga, A. (2024). A critical review towards artificial general intelligence: Challenges, ethical considerations, and the path forward. World Journal of Advanced Research and Reviews, 21(3), 1262-1268.
40. Van de Poel, I. (2020). Embedding values in artificial intelligence (AI) systems. Minds and machines, 30(3), 385-409.
41. Winfield, A. F., Michael, K., Pitt, J., & Evers, V. (2019). Machine ethics: The design and governance of ethical AI and autonomous systems [scanning the issue]. Proceedings of the IEEE, 107(3), 509-517.
42. Yang, Y., Ma, M., Huang, Y., Chai, H., Gong, C., Geng, H., ... & Wang, J. (2025). Agentic Web: Weaving the Next Web with AI Agents. arXiv preprint arXiv:2507.21206.
43. Yazdanpanah, V., Gerding, E. H., Stein, S., Dastani, M., Jonker, C. M., Norman, T. J., & Ramchurn, S. D. (2023). Reasoning about responsibility in autonomous systems: challenges and opportunities. Ai & Society, 38(4), 1453-1464.
44. Yuksel, K. A., & Sawaf, H. (2024). A multi-AI agent system for autonomous optimization of agentic AI solutions via iterative refinement and LLM-driven feedback loops. arXiv preprint arXiv:2412.17149.
45. Zuccotto, M., Castellini, A., Torre, D. L., Mola, L., & Farinelli, A. (2024). Reinforcement learning applications in environmental sustainability: a review. Artificial Intelligence Review, 57(4), 88.