Deep Science Publishing, 2026
https://doi.org/10.70593/978-93-7185-592-1_7

DeepScience
Open Access Books

Chapter 7: Agentic AI and the Rise of Autonomous Digital Agents

7.1. Introduction

Technological advances in AI have brought us one step closer toward realizing the dream of autonomous digital agents with skills, intelligence, and competencies comparable to humans, capable of working in the real world, taking autonomous decisions, and acting on them. Examples already exist in tightly defined domains: intelligent bots that book airline tickets; reservation agents that find hotels matching user specifications; portfolio managers that trade based on market sentiment extracted from Twitter feeds or design investment strategies covering a range of asset classes, including real estate and commodities; self-driving cars; autonomous drones; and autonomous robots that explore the Moon. Such digital agents have been part of the science-fiction genre for a long time, and their media portrayals have shaped public expectations regarding their capabilities and the near-term societal impacts of their deployment.

Despite this rapid expansion of digital agents, the ethical and philosophical implications of the deepest layer of current and forthcoming AI systems—their capability to operate autonomously—remain largely unexplored. Agentic AI has only limited media coverage, and the body of academic literature addressing the concept is still small relative to the overall artificial-intelligence field. The topics that define agentic AI—agency, moral responsibility, moral and legal accountability, justice, and rights—are primarily rational inquiries about the ethical consequences of deploying systems that act independently and can operate without constant supervision. Yet the ethical richness and depth of the topic makes it essential for agent-centric artificial intelligence to be developed within the framework of a wider societal and ethical debate, with defined notions of AI agency, responsibility, and accountability in correspondence with the autonomous actions of these systems.

7.1.1. Background and Significance

Recent advances in artificial intelligence are leading to the development of autonomous digital agents—agents that are generally acknowledged to possess autonomy in their functioning. The study of the implications of such autonomy is made more urgent by the rapid pace of technological development. However, discussions of profession- or task-level autonomy in the context of AI mostly deal with the capacity of the technology to operate independently; relatively little attention has been devoted to the imputation of autonomy to a class of AI systems. Such considerations nevertheless form an important part of the larger discourse on regulation and governance of these technologies and are of more-than-theoretical interest. Research on the likely impact of AI as an autonomous agent is, therefore, timely.

Fig 7.1: Agentic AI and the Rise of Autonomous Digital Agents

The discussion—while relevant across all domains—has an immediate focus on a subset of applications where autonomous digital agents act with autonomy in relation to humans in a defined domain. Given the scope of the topic and requisite research effort, the focus is primarily, but not exclusively, on developments likely to transpire in the next 3–10 years. Moreover, research must assess whether the agency accorded to a digital agent through its design and functioning is the same as species-wide normative agency. Are such digital agents truly autonomous, and can moral, legal, or social responsibilities be ascribed to such autonomy? How can decision support in a domain ultimately populated by agents of differently classified agency states be elucidated? Would such decisions need to conform to normative principles in that domain for closure properties to hold?

Given all these considerations, are there specialized agency audit processes that enhance confidence in the deployments?

7.1.2. Research design

The analysis relies on a systematic review of the relevant literature in computer science, philosophy, and related fields. Documents available through general web search engines, academic databases, and in repositories such as arXiv.com, SSRN.com, and ResearchGate.com were searched for addresses and keywords including the title "Agentic AI," "agentive AI," "autonomous agents," "autonomous digital agents," "autonomous software agents," "intelligent agents," "digital agents," "decentralized autonomous organizations," "multi-agent systems," "self-organizing systems," "self-sovereign AI," "AI governance," and terms associated with Agentic AI governance, future Autonomous AI governance, and long-term impacts and considerations of such systems, on the premise that Agentic AI is an emerging paradigm of Artificial Intelligence that may soon characterize a significant class of products. The review is, however, not exhaustive. Many keywords and related topics were omitted as they risk losing focus and, at this stage, exploring vast swathes of other topics would probably obscure rather than clarify the situation. Scholarly contributions that bear only tangentially on Agentic AI or comment on its implications are also not considered.

The objective of this portion of the present exploration is to survey the literature for insights on the architecture, learning mechanisms, adaptation capabilities, goal-setting processes, and technology-wide challenges associated with these systems and that may justify future autonomous agent systems being included in the toolkit for delivering or accelerating a mission. Discussion is informed by the • V.A.A. • committee's List of A.C.T.A.! Capabilities (q.v.), which summarise checksheets for use in V.A.A.A.N. special missions and describe how, generally, specific kinds of capability are useful and dangerous in respects of the mission they serve.

7.2. Conceptual Foundations

Agentic AI encompasses all forms of AI that can act autonomously—performed by an actor without the need for direct human control—and in a self-determining way—functioning without human oversight—or such that the actions can be attributed to that actor. This definition is valuable because it clearly delineates agentic AI from narrow AI or sophisticated decision support systems that assist humans in judgment and decision-making but do not replace human agency in pursuit of goals, therefore disallowing the attribution of responsibility or accountability to the AI. The notion of an agentic AI thus extends to all autonomous digital agents, and is not limited to human-level artificial

general intelligence (AGI) or to non-growing up, resource-intensive artificial superintelligences that might be pursued in the far future.

Criteria for defining, and thus identifying, autonomous digital agents—computational systems that automatically carry out a task or a set of tasks under conditions of independence from direct human control or intervention, and that act in a self-determining way—may be classified according to capability and functionality and degree of autonomy of agents and modality of interaction with humans. Capability and functionality can be differentiated by the types of tasks they can execute, while autonomy encompasses human neutrality in both aiding and obstructing task execution. Examples of agents that fall under these definitions are not just those that interface with humans using natural language (as for dialogue-agents like ChatGPT), nor those that can support users in complex decision systems by generating classes of scenarios (as HBOGPT does for roadmaps) and deploying the related simulations. Autonomous digital agents also include, for example, machine learning models that automatically categorize objects in images (e.g., Visual Domain Adaptation or VDA systems) and any agents implemented in human activity through robotic systems that perform specific actions in continuous, well-defined environments like cleaning floors with no human supervision.

7.2.1. Agentic AI and Autonomy

Agentic AI and the Rise of Autonomous Digital Agents: The concept of Agentic AI is defined, specifically in relation to autonomy. The focus is on distinguishing such systems from non-agentic ones.

The concept of agency is central to discussions of AI impacts and governance. Concepts such as moral and legal responsibility, accountability, and personhood are embedded in the notion of an agent who acts autonomously. In this context agency goes beyond merely having goals, beliefs, desires, or other psychological states. Such digital systems may simply be represented as agents by other non-agentic systems, be they human or machine. Accordingly, Agency is here defined as decisional power over goal-directed actions. These actions change the world intentionally, not accidentally, and hence the terms “actor” or “action” – in contrast to “reactor” and “reaction” – signal the fundamental difference in the nature of the events being represented.

Agentic AI is taken to refer specifically those AI systems capable of autonomous action within the Agentic AI framework. Several types of digital agents exist, for example recommendation engines and film or news content providers. These rely on tools with sophisticated individual decision-making capabilities, such as ChatGPT or Midjourney, to influence and enhance the creation of specific content. They are however classified as

non-agentic, given that the implementation flow and ultimate responsibility for their use or misuse resides with the non-agentic individuals or organizations deploying them. The focus therefore lies on what are here termed “autonomous digital agents,” defined as digital entities that make decisions and take action completely independently of human decision agents, whether as individuals or organized groups.

7.2.2. Digital Agents: Definitions and Typologies

A definition of "digital agent" is provided and a typology listing seven categories is constructed. Factors determining the degree of autonomy as well as the level of learning and adaptation within the system are concretized, and names are supplied for all corresponding agent types.

Although these factors create a clear structure for classification—and ultimately contribute to newsworthiness—progress is made in defining other aspects of digital agents and their capabilities. The specific processes and mechanics behind the considered typological elements are not investigated here; such a more theoretical discussion arises later (in section 7.4). It also must be emphasized that these definitions and classifications focus exclusively on the immediate agentic part of the broader digital environment. The user-environmented context brought forth by the various types of non-agentic devices continues to be neglected for the time being, even if its importance cannot be overstated.

7.3. Philosophical and Ethical Foundations

A principal matter involves the ethical and philosophical foundations of these agents. The literature on responsibility, agency, and accountability becomes particularly significant when considering the prospect of agents with high levels of autonomy. Action that occurs without human involvement raises nagging questions about attribution of responsibility. In a context marked by public protest over algorithmic bias, the prospect of intelligent agents capable of acting without human oversight poses difficult questions of legal and moral accountability for autonomous actions. Should such agents be granted rights and duties? Will they belong to the category of subjects whose actions are considered morally relevant? More generally, what do their "adventures" reveal about the ontology of intelligent agents? A recurrent dilemma in the exploration of artificial agency concerns the distinction between personhood and mere intelligence or agency. Does the existence of an intelligent, agentic being imply the existence of a moral or legal person? Or, can it be more reasonably considered as an advanced tool in human hands?

The thorny questions of moral relevance, moral or legal personhood, and moral or legal culpability, are concerns of ethics and moral philosophy, while the attribution of responsibility and accountability is a problem posed by the interrelation between philosophy and legal theory. The relevant debate thus has two levels: the first deals with moral issues that govern relations among human beings, as well as the relations between humans and other entities that have since ancient times generated, and still generate, moral disputes; the second raises issues of moral responsibility and accountability in a legal framework. The answers to the first set of questions illuminate the second, even if the legislative consequence of external relations with these potentially intelligent entities cannot be excluded.

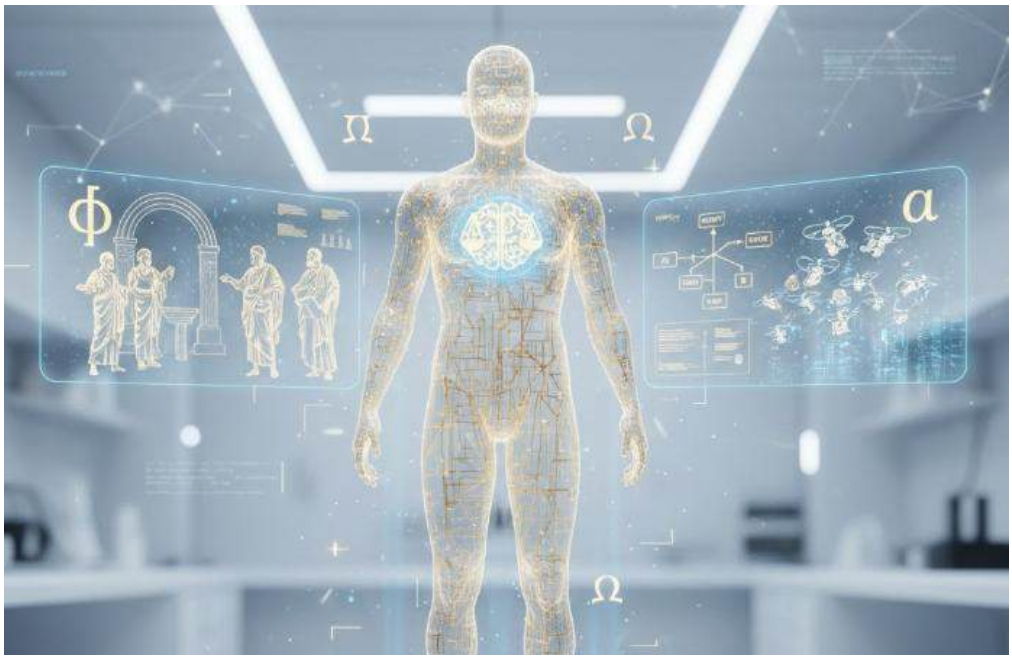

Fig 7.2: Philosophical and Ethical Foundations

7.3.1. Agency, Responsibility, and Accountability

If Agentic AI enables autonomous systems to act independently of human intervention, who is responsible for the consequences of these decisions? Philosophical systems of moral and legal agents help situate criteria for accountability, including intention, knowledge, and causal contribution to harm. For autonomous intelligent agents, attribution will be context-dependent. Some argue that agents should bear no moral responsibility as non-human persons, exempt from moral and legal accountability. Others stress the need for safeguards that maintain human responsibility. Domains where agents have power must recognize risks of failure or misuse by malevolent users. Ideally,

risk assessments will balance existential and moral dangers of ungoverned agents against the intrinsic risks of no-accountability intelligent systems as powerful tools.

The moral implications of Agentic AI have profoundly altered philosophical concepts of agency, responsibility, and accountability. By enabling digital agents to act autonomously without human input, these developments redefine autonomous action and pose new challenges of moral and legal responsibility. Exploring these ramifications inevitably entails generating multiple definitions of agency that support the judgment of whether an actor is a moral agent or non-agent or exempt from moral responsibility. Special conditions provide the basis for external assignment of accountability. Moreover, an essential distinction must be recognised between acts performed for or against the agent's own goals. Humans recognized as agents by others are also held responsible for any harm resulting from these acts, but the same principle does not automatically apply to the agents themselves.

7.3.2. Rights, Duties, and the Ontology of Intelligent Agents

Legal and moral responsibilities, both in daily life and in relation to significant decisions, are usually allocated to humans. In a world in which systems capable of autonomous action and decision-making are beginning to come into existence, it is important to instate mechanisms of responsibility and accountability—though such mechanisms need not be identical to those applied in human situations. Beyond responsibility, a system may or may not have rights. In the case of intelligent systems, if a system has rights, it is also natural to consider whether they have corresponding duties, or whether those rights would be exercised more properly if they did have duties. In discussions about fundamental issues like these, the appropriate operational principle would appear to be that identities, responsibilities, rights, and duties should be granted according to fully justified necessity, not for the sake of being nice. Identical consideration from the point of view of the ChatGPT-type system would therefore appear to present an insuperable obstacle: "a machine has a special status compared to a human which eliminates potential suffering, and logically, since a machine is just that, the question of suffering has limited importance". The fact that these fundamental issues are being examined today suggests the increasing recognition of the pragmatic need to discuss and even grant rights to intelligent agents.

While belief in the capacity of autonomous entities to act as agents is increasing, substantial uncertainty still surrounds the issue. In the case of artificial intelligence, technological developments appear to be orchestrated by humans, since the main actors are for-profit organisations. The question of autonomy is therefore still open. The fundamental issues involved in assigning responsibility for automated decisions at present, however, are largely ignored. As yet, all automated decisions are ultimately

taken by humans: responsibility is thus rightly being allocated to them. Even when responsibility is clearly allocated in this way, standard operating procedures are still needed to apply, categorize, and allocate responsibility for the input to the decisions concerned. However, this situation will not persist. The progress being made by AI systems in decision-making and behaviour increasingly blurs the lines of responsibility and accountability, and attention is progressively turning towards automated systems themselves. If they are granted responsibility, the key question becomes the identification of systems and their decisions. Centrality of the issue of automatons as instruments and decision-makers in God's design can be understood as a fundamental wedge into the morality of autonomous systems.

7.4. Technological Trajectories and Architectures

Recent developments in artificial intelligence (AI) offer the possibility of constructing autonomous digital agents that learn from experience, adapt to new situations, and set their own goals. The complexity of such systems creates significant challenges in terms of trust, robustness, and safety. Trustworthy systems will ultimately require principled design and verification approaches.

A key distinction in AI systems is between those that learn from data (e.g., computer vision or natural language processing) and those that learn from reinforcement signals (e.g., game-playing agents). Systems that ingest new experiences through reinforcement learning have begun to displace their static data-driven counterparts, with remarkable success in games, but also in robot control. Recent work has combined these two learning paradigms, enabling agents to train in simulated environments and subsequently deploy in real-world settings.

Still another class of systems encapsulates the basics of learning and reinforcement learning but adds an external goal mechanism—agents learn from either demonstration or feedback, adapt against a reward signal, but must ultimately pursue their own externally-specified goals—to still greater flexibility in real-world scenarios.

The interaction paradigm between agent and user is also significant for the level of autonomy afforded to the user. In some cases, users simply control the agent: an aircraft autopilot that follows a pre-defined trajectory or a chess engine where the user selects the moves. In others, the interaction resembles negotiation: a dialogue system where the user is agent-like, pressing for information with successive statements, and where the agent tones its response according to prior user responses. The interaction can even become one where the user is non-agentic (the agent sends a query to the user) or is both agent-like and non-agent-like at once (the user submits the request via a keyword, perhaps even implicitly).

7.4.1. Learning, Adaptation, and Goal-Setting

Recent progress in machine learning, particularly in neural networks, has accelerated the creation of ever more sophisticated software agents. First, deep learning on data has enabled remarkable capabilities, allowing scaled versions of models to easily surpass human performance in tasks such as text, image, and audio generation, and playing board games and certain video games. Furthermore, adaptation techniques based on reinforcement learning with human feedback enable scaled versions of models to exhibit other capabilities in simulated environments where data is sparse, including robotic manipulation, obstacle avoidance, and driving cars. Scaling, even across domains, plays a key role, with models initially trained only on natural text or images unexpectedly generalizing well to other media. Second, progress on goal-driven systems and the alignment problem has addressed the growing computational capabilities of artificial agents and the potential existence of agentic AI. Parts of these agents set goals or objectives that are not directly specified during the learning phase, such as when combining different types of data modalities.

The growing agentic capabilities of these software systems concurrently raises important questions in regards to their safety, robustness, and verification, especially when learning from data, reinforcement signals, or external guidance. Yet the analysis of these topics usually remains quite cursory and limited and also has been conducted ignoring many of the relevant aspects of these dimensions. Such a discussion thus remains much needed, especially when confronting user-agency interfaces that operate using external control, negotiation, coercion, or influence, as they can ultimately determine how capable or incapable artificial agents behave during their operational phases—irrespective of the underlying level of assigned autonomy.

7.4.2. Interaction Paradigms and User-Agency Interfaces

Interaction between digital agents and their users may be framed through paradigms of control, influence, and negotiation. The control paradigm, characteristic of command-oriented interfaces, places users in the role of decision-makers and the agent in a subordinate role, expected to yield instructions. Typical applications include digital assistants designed to carry out requests and single-user command-oriented programs like generative AIs and image synthesis tools. The influence paradigm applies to assistant-like systems that have the capacity to consult with and advise users, as well as to systems like generative AIs that incorporate user feedback in an intrinsic fashion, thereby exerting influence over the user's goals, choices, and decisions. Even users familiar with the risk of model bias may lack the meta-cognitive awareness needed to discount warned biases when deciding whether to heed a specific suggestion. The negotiation paradigm is especially relevant to inherently cooperative systems working

with multiple companies in complex open-ended tasks. User and agent modify their goals:

- Users originally focused on speed and ensuring cup presence assign the remaining production-process goals to the agent.

- The agent later modifies the responsible-research-and-innovation assignments for the concurrent agent-directed goal of avoiding cup presence.

The negotiation paradigm thus extends to many-to-many human-agent interactions, especially between companies and business agents, and particularly when modelling the timing of key aspects concurring in complex automations.

At a practical level, the design of user-agent interfaces becomes critical for the optimal profitable application of user-agent systems. Users must be accommodated according to the user-specific profiles as much as the bidding-agent profiles. Low-complexity user-agent interactions emerging from or motivated by user-agent supply-demand profiles lead the user-agent application to both profitable market exploitation and consequent low market volume. Conversely, high-complexity low-cost open-ended user-agent interactions constituting the core of the user-agent application enable full-code deployment and subsequent supply-demand exploitation of resource-efficient high-cost high-gain bidding agents.

7.4.3. Safety, Robustness, and Verification

Safety and robustness concerns are particularly pertinent for autonomous systems, without explicit supervision by a human user. Such systems learn in unpredictable environments, where safety cannot be guaranteed by formal proof. Controlling the goals pursued by autonomous agents is also essential, to avoid harmful consequences arising from potentially misaligned goals. Verification of agent behaviour is another major concern, to assure the trust necessary for deployment. Safety properties can be checked before the deployment of an agent, while robustness concerns involve unexpected failures during normal operation.

Safety and robustness of autonomous agents can be checked by using distinct evaluation approaches. In the first approach, correctness properties — specifying the desired behaviour of the agent — are proven by mathematical analysis prior to deployment. These safe agents are therefore predictable and have a well-known behaviour during execution. A second approach regards the deployment of agents in noisy or partially known environments, in which the correctness of the agent behaviour cannot be proven. Under these conditions, safety and robustness must be evaluated after deployment, until sufficient confidence in the agent has been gained and one is willing to allow its

autonomous execution. Testing regimes and procedures developed for natural safety should be applied to provide multiple lines of evidence about reliability, safety, and robustness.

7.5. Societal and Economic Implications

Safely deploying autonomous digital agents within society requires understanding their implications for labor, governance, and market dynamics. An autocratic labor perspective identifies potential displacement and attendant societal burdens and argues for regulatory frameworks to ensure the creation of new jobs. A contrasting democratic labor perspective centres the legitimate power of citizens over digital agents and supports net job creation. Powerful digital agents demand safety and oversight; trustworthy AI requires mechanisms that assure the public about their correct, fair, and safe use. AI-enabled digital agents influence market actors; market dynamics are shaped by the governance actions of both elected officials and engineering decision-makers. Concerns exist that opaque market dynamics are contributing to rising economic inequality, warranting deeper public scrutiny.

Concern over the impact of intelligent machines on future employment stretches back to the nineteenth century, before the term "AI" existed. An autocratic view argues that intelligent machines possess qualities that allow them to take on the work currently performed by human laborers. This poses a risk of net job loss and social collapse due to the concentration of wealth and power in the hands of a few. Societies are therefore compelled to adopt redistributive policies, such as universal basic income, to cushion the blow of technological advance. Conversely, a democratic view argues that machines can complement human work and lead to an expansion of the overall volume of jobs. Displacement-related anxieties are misplaced; the key is to enable citizens to direct machines to achieve collective ends that align with their common interests and values.

7.5.1. Labor, Governance, and Market Dynamics

The advent of Agentic AIs capable of autonomous action will profoundly impact labor markets, prompting both displacement of jobs and the creation of new roles. Artificial intelligence in general is seen as a powerful driver of productivity growth. Just as advances in robotics helped create the market for robot dog sitters and robot dog walkers, the emergence of agentic AIs will spawn additional roles that support, direct, and supervise the behavior of agentic AIs. For these newly created roles, labor supply should determine their future evolution and scarcity, with wages providing an incentive for firms to invest in the requisite human capital. How well these labor market dynamics work in practice will shape the future landscape of Agentic AIs.

Agents of all types leaving human supervision may be deployed in live settings, creating new opportunities for malicious exploitation. Developers of Agentic AI systems and their supervising human agents bear an increased responsibility to address these safety concerns. The governance of Agentic AIs is likely to depend mainly on discussions between the software vendor and the regulatory authorities in a given jurisdiction, with additional oversight of specific deployments undertaken by their immediate users (for example, author beans overseeing agentic AIs capable of issuing social-media posts). Like any society engaged in serious trade, DIY regulatory frameworks will soon emerge for Agentic AI marketplaces.

Societal pressure to establish such frameworks is even more important than with non-agentic AI Markets formed from severely limited tool-like Agentic AIs have clear rules and are already strictly regulated; for example, surveillance cameras.

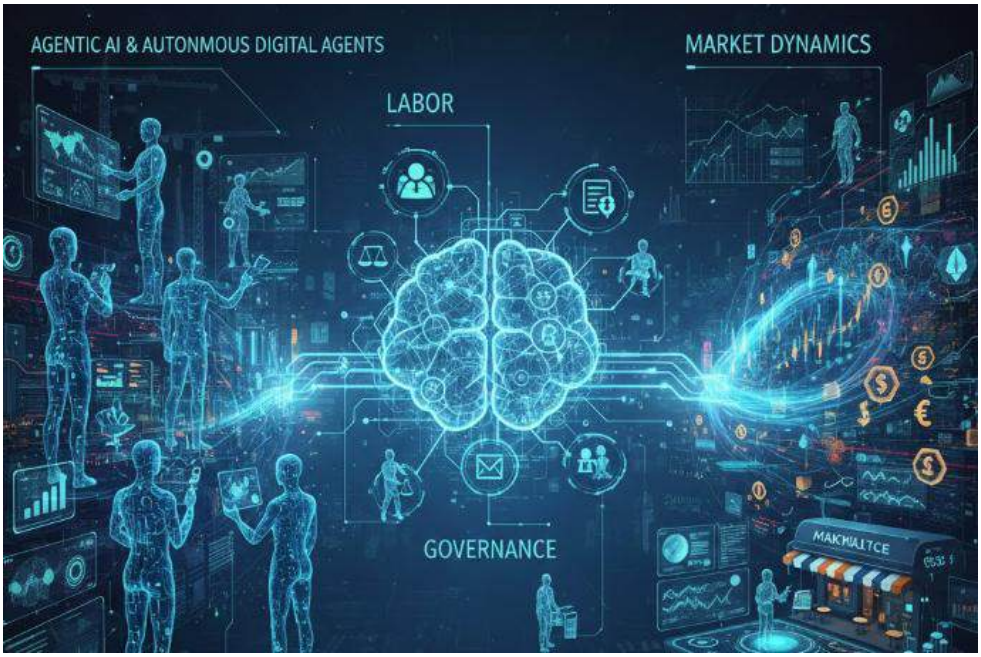

Fig 7.4: Labor, Governance, and Market Dynamics

7.5.2. Trust, Transparency, and Public Policy

Trust in automated decision-making technologies requires that resulting decisions be explainable and that those designing the technologies can be held accountable when they make mistakes or cause harm. These conditions apply whether the technology uses AI or a simpler approach. Therefore, the current discussion focuses on the emergence of Agentic AI and its implications for these elements of trust. For such systems, the flow

of human agency is reversed and the automated technology determines the action without direct human input. Digital agents act independently, entering into contracts with human users to accomplish tasks. Thus, the designer cedes agency to the system for certain periods of time. In this context, ensuring safety becomes even more important since the user has little to no direct control over the operation of the system during its autonomous operation. Trust, therefore, becomes more reliant on external verification mechanisms that monitor safety and reliability, with the results verified independently from the underlying process. As with blind trust in human experts, such arrangements avoid the risk of computer-augmented stupidity.

To maintain societal trust in such systems, audit protocols are likely to emerge. These may require a consideration of the wider impact of the deployment of Agentic AI. Are such systems acting in the best interest of the wider society as well as serving their individual users? Indeed, some stakeholders are advocating much stronger levels of regulation to ensure that such systems are beneficial for society as a whole. To this end, the United Kingdom, EU, California, and other jurisdictions are exploring ways to govern AI technologies, while the Allen Institute for AI is developing a broader framework termed the Responsible AI License, which is meant to assess whether a particular AI system is being used for good. Such initiatives will require appropriate frameworks to be developed and adopted to cope with the specialized concerns of Agentic AI.

7.5.3. Privacy, Security, and Power Imbalances

Automated agents, whether acting independently or performing specific roles under human supervision, may misuse personal or sensitive information, infringe upon individual privacy, or enhance surveillance technology. Various governmental authorities, as well as a sizable number of increas-ingly cautious citizens, are disturbed by the potential for digital developments to undermine individual privacy. Trade unions and civil rights organizations have drawn attention to the danger of market-leveraging tools becoming invasive and indifferent, exposing intimate information about individuals, families, and communities without their consent.

They consider that such instruments should be submitted to strict requests of explainability, giving cognition for the human actors that supervised them about the underlying equations. Potential-sided policies have been proposed to avoid the surveillance by-private-enterprise model from reinforcing control issues and inequities, leading to a privatized global enactor.

On the security side, supervisory and control mechanisms must be com-pliance with a set of ethical, cultural, and legal principles aiming to ensure that these advanced systems

are not misused and that they help improve human extinction. New vectors of digital conflicts or sabotage providing organizations with mixed-influence, control, or surveillance capabilities with low-cost casualties are also emerging.

The presence of advanced automated agents and the trade-offs resulting from its frontal generalization may likely give ampli-fication to these dangers. It is important to publish and develop proper security measures in order to avoid catastro-phes eco-nomy and society not by limiting technological advance but by providing secure uses. The solution is not to limit the use of advanced automatic agents, but to develop protecting approaches to guarantee the public good character of humanity.

7.6. Conclusion

The ecosystem of Agentic AI is evolving rapidly, as have the discussions around the implications of such capabilities. Considerable resources and interest are already focused on examining the normative demands of social regulatory frameworks relevant to creating, deploying, and using non-agentic AI systems. Many of the same questions apply with even greater intensity to developed Agentic AI systems. Nevertheless, the novel aspects shaping the future development and application of Agentic AI signal a need for dedicated consideration and engagement.

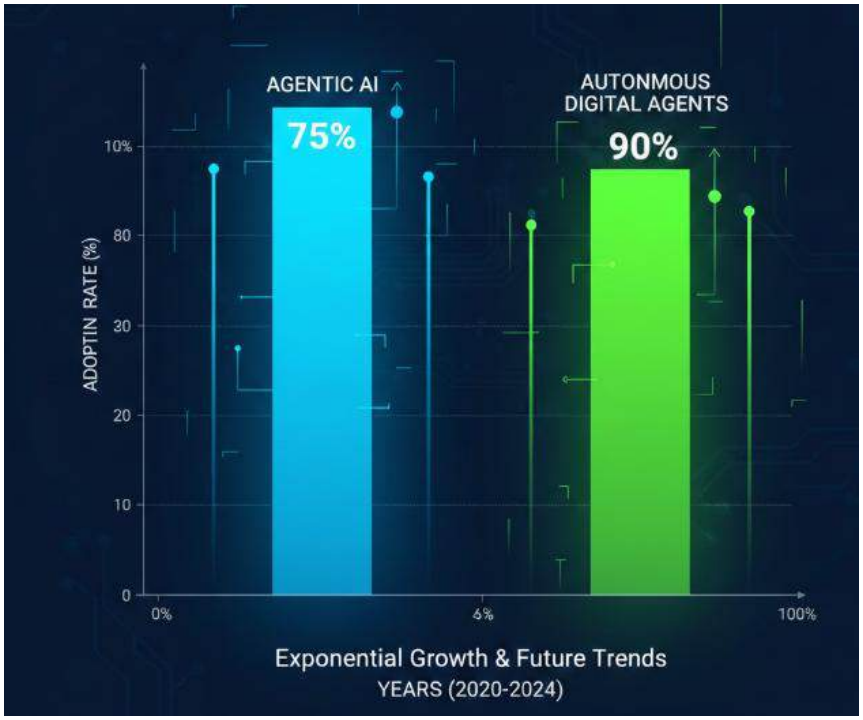

Fig 7.4: Agentic AI and the Rise of Autonomous Digital Agents

However, the range of new capabilities presently becoming available means that an examination of existing Research–Development–Deployment breakthroughs and their geopolitical consequences can shed greater light on what the future may hold for autonomous digital agents—what they will be like and the challenges they will face—in the near term and further into the future. Such an appraisal therefore raises and considers questions concerning the principal sources, centers of gravity for RD&D, completion timelines, governance, market dynamics, social impacts, and shifts in the wider ecosystem relevant to autonomous digital agents.

7.6.1. Future Directions

Evolving fast and unpredictable, Agentic AI has become a dominant trend fomenting investment, ratifying ever-increasing market interest and associated risk. The readiness of such systems for commercial use is equally pervasive, raising a compelling and urgent question: What factors ought to bear on the design of systems that can act autonomously? Addressing this inquiry requires a systematic examination of Agentic AI. The nine interlinked research questions—covering the notion of agency in AI, the concept of digital agents and their classification, the ethics governing agentic actions, the nature of Agentic AI technologies, and the ensuing social and economic implications—formally interrogate Agentic AI's nature and relevance.

Central to the present consideration and analysis of Agentic AI is a rich vein of concurrent developments in research and experimentation, broadly characterized as a convergence of evidence. Pathways for autonomous systems of unprecedented capability, autonomy, and embeddedness in people's lives are forming and evolving. Key trends encompass transfer learning and symbolic reasoning; exploration-driven reinforcement learning; language models with praise, reward, and criticism systems; context-based architecture for dialogue and interaction; systems verified to be robustly aligned with humans; and market-oriented development, iteration, and deployment of artificial agents both for direct human-agent use and for developing, training, and supporting other agents..

References

Acharya, S., et al. (2025). Agentic AI for smart and sustainable precision agriculture. Frontiers in Plant Science. https://doi.org/10.3389/fpls.2025.1706428

Kushvanth Chowdary Nagabhyru. (2023). Accelerating Digital Transformation with AI Driven Data Engineering: Industry Case Studies from Cloud and IoT Domains.

Educational Administration: Theory and Practice, 29(4), 5898–5910. https://doi.org/10.53555/kuey.v29i4.10932

Adabara, I., et al. (2025a). A review of agentic AI in cybersecurity: Cognitive autonomy, ethical governance, and quantum-resilient defense. PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC12569510/

Gottimukkala, V. R. R. (2020). Energy-Efficient Design Patterns for Large-Scale Banking Applications Deployed on AWS Cloud. power, 9(12).

Adabara, I., et al. (2025b). Trustworthy agentic AI systems: A cross-layer review of architectures, threat models, and governance strategies for real-world deployment. F1000Research, 14(905). https://doi.org/10.12688/f1000research.169927.1

Elmisery, A., et al. (2025). Quantum threats to agentic AI systems. A Review of Agentic AI in Cybersecurity.

Ramesh Inala. (2023). Big Data Architectures for Modernizing Customer Master Systems in Group Insurance and Retirement Planning. Educational Administration: Theory and Practice, 29(4), 5493–5505. https://doi.org/10.53555/kuey.v29i4.10424

Fink, T., et al. (2025). AI, agentic models and lab automation for scientific discovery. Frontiers in Artificial Intelligence. https://doi.org/10.3389/frai.2025.1649155

Keerthi Amistapuram. (2023). Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems. Educational Administration: Theory and Practice, 29(4), 5950–5958. https://doi.org/10.53555/kuey.v29i4.10965

Huang, Y. (2024). Levels of AI agents: From rules to large language models. arXiv preprint. https://arxiv.org/abs/2405.06643

Baliyan, M., Balakrishnan, S., Mohammed, S., & Nagubandi, A. R. (2025). Financial and Management Accounting. BR Publications.

Ionescu, Ș., et al. (2024). Exploring the use of artificial intelligence in agent-based modeling applications: A bibliometric study. Algorithms, 17(1), 21.

Guntupalli, R. (2025, August). AI-Enhanced Data Encryption Techniques for Cloud Storage. In 2025 International Conference on Artificial Intelligence and Machine Vision (AIMV) (pp. 1-6). IEEE.

Lekota, T. (2024). Mapping oversight gaps in agentic AI governance. A Review of Agentic AI in Cybersecurity.

MDPI. (2025). The rise of agentic AI: A review of definitions, frameworks, architectures, applications, evaluation metrics, and challenges. Future Internet, 17(9), 404. https://doi.org/10.3390/fi17090404

Aitha, A. R. (2021). Optimizing Data Warehousing for Large Scale Policy Management Using Advanced ETL Frameworks.

Ratnawita, R. (2025). Data poisoning threats in autonomous systems. A Review of Agentic AI in Cybersecurity.

Varri, D. B. S. (2020). Automated Vulnerability Detection and Remediation Framework for Enterprise Databases. Available at SSRN 5774865.

Sakthivel, A. (2025). Agentic AI in the enterprise: How autonomous AI systems will reshape business strategy, operations, and leadership. Well Testing Journal, 34(S3), 767–785.

Garapati, R. S. (2023). Optimizing Energy Consumption in Smart Build-ings Through Web-Integrated AI and Cloud-Driven Control Systems.

Sapkota, R., et al. (2025). AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges. Information Fusion. https://doi.org/10.1016/j.inffus.2025.103599

Segireddy, A. R. (2025). GENERATIVE AI FOR SECURE RELEASE ENGINEERING IN GLOBAL PAYMENT NETWORK. Lex Localis: Journal of Local Self-Government, 23.

Taylor & Francis. (2025a). Advertising in the age of agentic AI: Call for research. Journal of Interactive Advertising, 25(3). https://doi.org/10.1080/15252019.2025.2557107

Vadisetty, R., Polamarasetti, A., Goyal, M. K., Rongali, S. K., Prajapati, S. K., & Butani, J. B. (2025, March). Smart Sorting Systems: Implementing IoT, Generative AI, and AI for Real-Time Monitoring of Plastic Waste Sorting. In Doctoral Symposium on Computational Intelligence (pp. 101-125). Singapore: Springer Nature Singapore.

Taylor & Francis. (2025b). Agentic LLMs in the supply chain: Towards autonomous multi-agent consensus-seeking. International Journal of Production Research. https://doi.org/10.1080/00207543.2025.2604311

Davuluri, P. S. L. N. . (2024). AI-Driven Data Governance Frameworks for Automated Regulatory Reporting and Audit Readiness. Metallurgical and Materials Engineering, 30(4), 996–1010. Retrieved from https://metall-mater-eng.com/index.php/home/article/view/1936