The Rise of Autonomous AI Agents: Automating Complex Tasks

Michael Negnevitsky[1,*]

[1]Oxford University, Oxford OX1 2JD, UK
Corresponding author: Michael Negnevitsky (e-mail: michaelnegnevitsky@gmail.com).

DOI: https://doi.org/10.63619/ijai4s.v1i2.007

Manuscript received May 30, 2025; revised June 13, 2025; published July 11, 2025.

Abstract: The emergence of autonomous AI agents represents a transformative leap in the evolution of artificial intelligence. These intelligent systems, capable of independently perceiving environments, making decisions, learning from experience, and executing multi-step actions without continuous human oversight, are redefining the boundaries of what machines can accomplish. Unlike traditional rule-based or supervised AI systems, autonomous agents integrate deep learning, reinforcement learning, natural language processing, and multi-modal decision frameworks to solve complex, dynamic, and often ambiguous real-world problems. This paper explores the technological underpinnings, capabilities, applications, and implications of autonomous AI agents. It critically examines their deployment in sectors such as healthcare, finance, cybersecurity, logistics, manufacturing, education, and scientific research. Furthermore, it addresses the ethical, legal, and socio-technical challenges arising from the increasing autonomy of machines, offering a roadmap for responsible innovation. Ultimately, autonomous AI agents are not merely tools—they are collaborators in a new era of intelligent automation.

Keywords: Autonomous AI agents, intelligent automation, reinforcement learning, multi-agent systems, task automation, artificial general intelligence, ethical AI, autonomous decision-making, AI planning, agent-based modeling.

1. Introduction

The 21st century has witnessed unprecedented advancements in artificial intelligence (AI), transforming industries, economies, and daily life. Among these innovations, the emergence of *autonomous AI agents* marks a paradigm shift—not merely in computational capability, but in the delegation of complex cognitive tasks to machines [1], [2]. These agents, which can independently perceive environments, make context-aware decisions, learn from experience, and execute goal-directed actions, are rapidly redefining what constitutes automation in the modern world [3].

Unlike traditional narrow AI systems that operate within static, rule-based frameworks or require continuous human oversight, autonomous agents exhibit a high degree of autonomy, adaptability, and generalization. They are capable of real-time reasoning, dynamic planning, and lifelong learning in open-ended, unpredictable environments [4], [5], [6]. For instance, self-driving vehicles navigate chaotic traffic, robotic surgeons make intra-operative decisions, and language agents engage in complex, multi-turn dialogues—all with minimal or no human intervention [7], [8]. These developments represent not just engineering milestones, but the initial foundations of artificial general intelligence (AGI)—a form of intelligence that can flexibly perform a wide range of cognitive tasks across domains [9], [10], [11].

The rise of autonomous agents also reflects a broader convergence of AI subfields, including deep reinforcement learning, multi-agent systems, neuro-symbolic reasoning, and large language models [12]. These integrations have enabled agents to handle not only physical tasks in robotics and logistics, but also

abstract reasoning tasks such as legal document drafting, scientific hypothesis generation, and adaptive education delivery.

However, this technological leap also brings profound societal and ethical challenges [13]. Delegating decisions to non-human entities raises critical concerns: How do autonomous agents learn, adapt, and make decisions in high-stakes environments? What domains are they most suited for—and where should human control remain central? How can we ensure these agents behave in ways aligned with human values, especially when their actions affect safety, justice, or equity? What regulatory, technical, and governance frameworks are required to manage the deployment of such intelligent systems?

This paper offers a comprehensive examination of the architecture, applications, training methodology, and ethical implications of autonomous AI agents. Through illustrative use cases, comparative analyses, and future outlooks, we aim to understand not only what these systems can do, but also what they *should* do—as intelligent collaborators in a world increasingly shaped by machine agency.

2. The Architecture of Autonomous Agents

Autonomous agents are intelligent systems capable of perceiving their environment, making decisions, taking actions, and adapting over time without continuous human intervention [14], [15]. Their architecture is typically organized into modular and hierarchical layers, each responsible for distinct aspects of functionality [16], [17]. This layered approach enhances interpretability, modular development, and scalability. We describe the four primary layers of an autonomous agent system: perception, cognition, action, and memory/adaptation [18], [19].

2.1. Perception Layer

The perception layer serves as the sensory interface between the agent and its environment. It transforms raw data into structured representations that higher-level modules can interpret and reason over [20], [21].

- **Computer Vision:** Enables the agent to understand visual input, including object detection, scene segmentation, motion tracking, and spatial layout analysis. For example, a drone may identify roads, humans, or wildlife using YOLO or Mask R-CNN.
- **Natural Language Processing (NLP):** Allows the agent to interpret textual or spoken instructions, conduct dialogue, and extract semantic meaning. Applications include language-guided navigation and collaborative task execution with humans.
- **Sensor Fusion:** Combines data from multiple modalities—e.g., LiDAR, RGB cameras, thermal sensors, radar, and microphones—to build a robust and redundant perception system that improves accuracy in uncertain environments.

This layer ensures the agent has a coherent, real-time understanding of its surroundings.

2.2. Cognitive Layer

The cognitive layer is the "brain" of the agent. It interprets sensory inputs, generates internal goals, reasons about consequences, and chooses actions [22], [23].

- **Reinforcement Learning (RL):** Enables agents to learn optimal policies through trial-and-error interaction with the environment. This is widely used in autonomous driving, game-playing, and robotic control.
- **Meta-learning:** Also known as "learning to learn," this allows agents to rapidly adapt to new tasks or environments with minimal data, enhancing their generalization capabilities.
- **Planning and Scheduling:** Classical AI techniques such as A*, Monte Carlo Tree Search (MCTS), or PDDL-based planners are used to generate multi-step action plans under constraints.
- **Neural-Symbolic Integration:** Combines neural networks (for perception and learning) with symbolic reasoning (e.g., logic rules, knowledge graphs) to achieve both flexibility and interpretability.

Together, these techniques enable the agent to make informed, strategic decisions in dynamic environments.

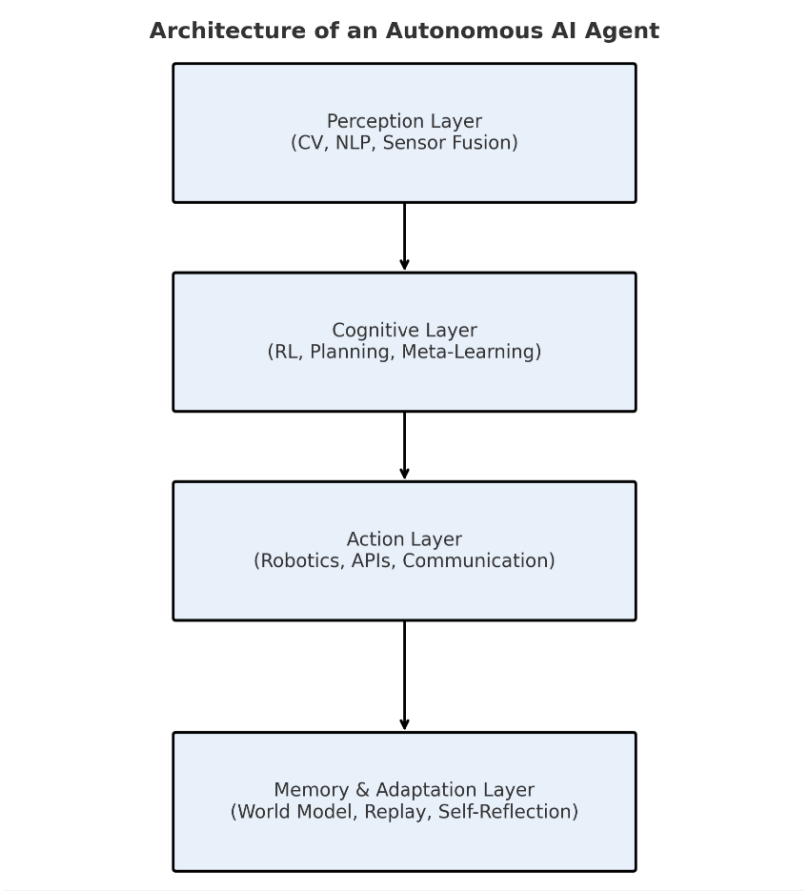

Fig. 1. Layered architecture of an autonomous AI agent system, illustrating perception, cognition, action, and memory components.

2.3. Action Layer

Once a decision has been made, the action layer is responsible for physically or virtually executing that decision [24], [25]. It acts as the interface between cognition and the external world.

- **Robotic Actuation:** In embodied agents, this involves motor commands to manipulators, drones, or vehicles, enabling locomotion, manipulation, and interaction with objects.
- **API-based Execution:** In software agents (e.g., trading bots, digital assistants), this may involve API calls, web automation, or remote database queries.
- **Multi-agent Communication:** For agents operating in teams or swarms, this includes protocols for coordination, negotiation, and consensus (e.g., using ROS, MQTT, or custom messaging layers).

This layer ensures the agent can carry out tasks in the physical or digital realm.

2.4. Memory and Adaptation Layer

This layer equips agents with persistence, self-awareness, and the ability to evolve [26], [27].

- **World Models:** Agents maintain an internal representation of the environment (spatial maps, object models, social dynamics), which is updated over time based on perception and outcomes.
- **Experience Replay and Logging:** Historical data, successes, and failures are stored and sampled for continual learning or offline optimization.
- **Self-reflection and Adaptation:** More advanced agents incorporate introspection to revise strategies,

detect anomalies, or alter behaviors in novel or adversarial contexts.

This layer is crucial for long-term autonomy, especially in open-world settings where change is constant [14], [28]. In summary, this multi-layer architecture supports a full autonomy loop—from sensing and interpreting the environment, to making and executing decisions, to adapting and improving over time [29], [30], [31]. Each layer builds upon the outputs of the previous one, enabling robust and generalizable AI agents across domains including robotics, virtual assistants, autonomous vehicles, and scientific discovery [32], [33].

3. Applications of Autonomous AI Agents

Autonomous AI agents are increasingly deployed in a wide range of high-impact domains, where their ability to perceive, reason, act, and adapt brings measurable improvements in efficiency, accuracy, and scalability. This section outlines key areas of application [34], [35]:

3.1. Healthcare

- **Clinical Assistants:** Autonomous diagnostic agents that analyze patient data, suggest tests, or offer differential diagnoses [36], [37].
- **Surgical Robots:** AI-driven systems capable of making fine-grained decisions during surgery, adapting to unexpected complications in real time [38], [39].
- **Virtual Therapists:** NLP-enabled agents that provide cognitive behavioral therapy (CBT), personalized to patient history and engagement style [40].

3.2. Finance

- **Autonomous Trading Agents:** Deep reinforcement learning agents that identify and exploit temporal patterns in financial markets [41].
- **Robo-Advisors:** Automated systems offering personalized investment strategies and dynamic portfolio rebalancing [42].
- **Fraud Detection:** Agents that monitor transactions in real time to flag anomalous patterns and adapt to new fraud tactics [43].

3.3. Manufacturing and Industry 4.0

- **Smart Factory Bots:** Autonomous robots that manage inventory, collaborate across supply chains, and self-optimize workflows.
- **Predictive Maintenance:** Agents that analyze sensor streams to anticipate equipment failures and schedule maintenance preemptively.

3.4. Logistics and Transportation

- **Autonomous Vehicles:** Delivery drones, autonomous trucks, and warehouse robots for end-to-end logistics automation.
- **AI Dispatch Systems:** Intelligent agents that optimize fleet routing, reduce idle time, and respond to demand shifts dynamically.

3.5. Cybersecurity

- **Network Defense Agents:** Autonomous systems that patrol networks, detect intrusions, and initiate automated countermeasures.
- **Adversarial Agents:** Simulated attackers used to probe system vulnerabilities and test cyber-defense robustness.

3.6. Scientific Discovery

- **Autonomous Laboratory Agents:** Robotic platforms that design hypotheses, run experiments, and analyze results with minimal human input.
- **Applications:** Drug discovery, protein structure prediction (e.g., AlphaFold), and materials design.

3.7. Education and Learning

- **Intelligent Tutoring Systems:** AI agents that adapt instructional content to individual student performance and learning styles.
- **AI Mentors:** Simulations that expose learners to real-world challenges and guide them through problem-solving exercises.

TABLE I
SUMMARY OF APPLICATION DOMAINS FOR AUTONOMOUS AI AGENTS

Domain	Key Applications	Agent Capabilities
Healthcare	Diagnostic assistants, surgical robots, virtual therapists	Clinical reasoning, real-time decision-making, dialogue personalization
Finance	Trading bots, robo-advisors, fraud detection	Market adaptation, risk assessment, anomaly detection
Manufacturing	Smart factory bots, predictive maintenance	Multi-agent coordination, sensor-based prediction
Logistics	Delivery drones, AI fleet dispatch	Route optimization, dynamic scheduling
Cybersecurity	Threat detection, adversarial simulation	Network monitoring, real-time response, self-defense
Scientific Discovery	Automated labs, drug discovery, protein folding	Hypothesis generation, experiment design, model-driven exploration
Education	Tutoring systems, AI mentors	Adaptive learning, scenario simulation, personalized feedback

4. Methodology of Agent Training and Deployment

The development pipeline for autonomous AI agents involves a sequence of well-structured stages [44]. Each stage is critical to ensuring that agents learn effectively, generalize well across environments, and perform safely and reliably in real-world applications [45]. This section outlines the typical training-to-deployment workflow.

4.1. Task Definition and Environment Design

The first step in developing an autonomous agent is to define the task specifications [46]. This includes the objective function, success criteria, environmental dynamics, and constraints such as time limits, safety rules, or energy budgets [47].

- **Environment Setup:** Training begins in controlled, simulated environments such as OpenAI Gym, MuJoCo, Habitat AI, or Isaac Sim.
- **Reward Shaping:** Proper design of reward functions is essential to ensure that the agent learns desired behaviors without unintended side effects.
- **Curriculum Learning:** Environments can be progressively scaled in complexity, allowing agents to acquire skills in stages.

Simulators offer safe, fast, and cost-effective platforms for early development and benchmarking.

4.2. Reinforcement Learning and Policy Optimization

Agents learn to map observations to actions by maximizing cumulative rewards. This stage uses reinforcement learning (RL) algorithms to iteratively improve the policy [48], [49].

- **Deep Q-Networks (DQN):** Value-based learning for discrete action spaces, especially effective in game-like scenarios.
- **Proximal Policy Optimization (PPO):** A policy-gradient method that balances stability and sample efficiency, widely used in continuous control tasks.
- **Multi-agent RL (MARL):** Enables training of agents in competitive or cooperative environments with other autonomous agents.

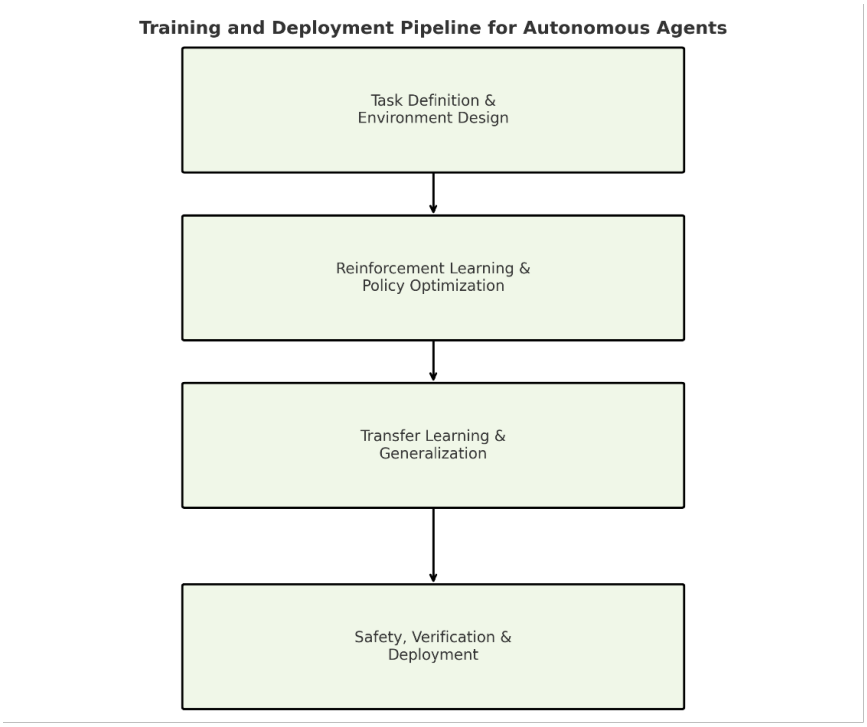

Fig. 2. Training and deployment pipeline of autonomous AI agents, from task specification to safe real-world execution.

Training often involves millions of episodes, parallelized rollouts, and GPU-accelerated optimization.

TABLE II
COMPARISON OF REINFORCEMENT LEARNING ALGORITHMS FOR AUTONOMOUS AGENTS

Algorithm	Type	Strengths / Suitable Scenarios
Deep Q-Network (DQN)	Value-based	Effective in discrete action spaces, e.g., game environments
Proximal Policy Optimization (PPO)	Policy-gradient	Stable and sample-efficient, widely used in continuous control
Multi-Agent RL (MARL)	Multi-agent	Supports cooperation and competition between multiple agents
A3C / A2C	Asynchronous Policy-based	Fast convergence in large-scale simulation, suitable for parallelized training

4.3. Transfer Learning and Generalization

One of the main challenges in deploying autonomous agents is bridging the gap between simulation and the real world [50], [51].

- **Domain Randomization:** Injects variability into simulations (e.g., lighting, textures, physics) to improve generalization.
- **Sim-to-Real Adaptation:** Techniques such as fine-tuning, adversarial domain adaptation, or representation disentanglement help transition to real-world deployment.
- **Continual Learning:** Architectures that support incremental learning prevent catastrophic forgetting and allow agents to update their knowledge over time.

These techniques ensure robustness under distributional shifts and enable long-term adaptability.

4.4. Safety, Verification, and Testing

Before deployment, autonomous agents must undergo rigorous safety evaluation and reliability testing [52], [53].

- **Formal Verification:** Mathematical proofs or symbolic model checking can guarantee properties such as reachability, safety bounds, or deadlock freedom.
- **Human-in-the-Loop Simulation:** Agents are tested with simulated or real human collaborators or supervisors to ensure behavior alignment.
- **Adversarial Testing:** Agents are exposed to edge cases, perturbations, or adversarial attacks to uncover hidden failure modes.
- **Shadow Deployment:** Agents operate in parallel with human operators or baselines in real settings, without direct control, to gather performance data before activation.

Safety is not a final step, but a continual process, monitored and refined post-deployment via feedback loops.

5. Ethical and Societal Implications

With power comes responsibility—autonomous AI agents, while promising unprecedented gains in efficiency and intelligence, also introduce complex ethical and societal challenges. These concerns must be addressed not only through technical safeguards, but also through transparent governance and inclusive stakeholder engagement [54], [55].

5.1. Decision Accountability

A fundamental question arises: Who is accountable when an autonomous agent makes a harmful or unlawful decision? This dilemma becomes particularly urgent in contexts such as autonomous vehicles causing accidents or medical AI agents misdiagnosing patients [56].

- Should responsibility lie with the original developers, the system deployers, or the organization that relies on the agent's outputs?
- Current legal systems struggle to handle such "algorithmic opacity," leading to calls for auditable AI and explainable decision pipelines.

Emerging proposals such as algorithmic impact assessments and liability insurance for AI are gaining traction.

5.2. Bias and Discrimination

AI agents trained on historical or skewed datasets risk perpetuating or even amplifying social biases. This can result in:

- Discriminatory hiring bots
- Biased medical triage algorithms
- Unequal resource allocation in public services

To mitigate this, fairness-aware machine learning and bias detection tools must be embedded into the training pipeline [57]. Techniques such as re-weighting, adversarial debiasing, and counterfactual analysis are increasingly used in agent design.

5.3. Autonomy vs. Human Control

While autonomy is the goal of intelligent agents, unchecked autonomy can lead to safety and ethical failures.

- In high-stakes domains such as defense, autonomous weapon systems raise existential concerns.
- In healthcare, a balance must be struck between automated recommendations and human clinical judgment.

Approaches such as human-in-the-loop (HITL), human-on-the-loop (HOTL), and adjustable autonomy architectures provide graded control.

5.4. Labor Displacement

Autonomous agents are expected to replace not only manual labor but also knowledge work in fields like legal analysis, journalism, and education [58].

- This technological unemployment may disproportionately affect low- and middle-skilled workers.
- It raises long-term questions about social equity, universal basic income (UBI), and the future of human labor.

Policies focusing on workforce retraining, lifelong learning, and equitable AI access are essential to mitigate harm.

5.5. Security and Manipulation

As autonomous agents become more capable, they also become more vulnerable to misuse and adversarial exploitation [59].

- Agents can be tricked with adversarial examples—e.g., images or commands that fool vision or language models.
- Social engineering or sensor spoofing can hijack autonomous systems for malicious purposes.
- Autonomous misinformation bots and large-scale behavioral manipulation are growing concerns.

Defensive techniques—such as robust training, adversarial testing, and secure architecture design—must become standard practice.

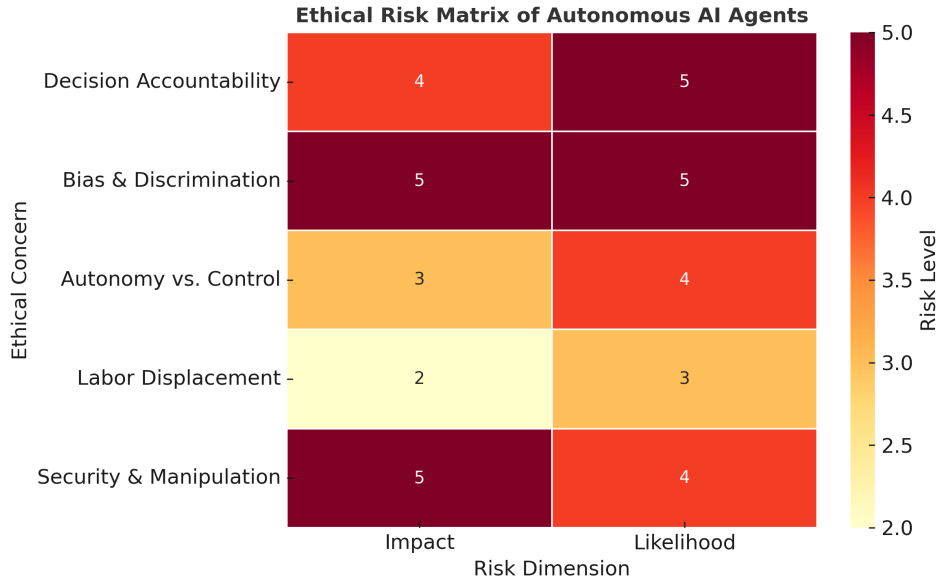

Fig. 3. Ethical risk matrix for autonomous AI agents, comparing potential impact and likelihood across key societal concerns.

6. Comparative Analysis of Traditional Systems vs. Autonomous AI Agents

Autonomous AI agents differ significantly from traditional software and automation systems in their learning ability, adaptability, and real-time decision-making capacity [60], [61]. Table III summarizes these contrasts across multiple domains.

TABLE III
COMPARISON BETWEEN TRADITIONAL SYSTEMS AND AUTONOMOUS AI AGENTS

Domain	Traditional Systems	Autonomous AI Agents
Healthcare	Rule-based diagnosis tools	Self-learning diagnostic agents that improve from new data and feedback
Finance	Scripted trading algorithms	Adaptive, self-optimizing trading bots that react to market volatility
Manufacturing	PLC-driven assembly robots	Multi-agent systems coordinating production lines with dynamic rescheduling
Cybersecurity	Signature-based threat detection	Real-time adaptive threat response agents capable of anomaly detection
Education	Static e-learning modules	Interactive, personalized learning tutors that adapt to student needs

7. Future Outlook: Towards Artificial General Intelligence (AGI)?

Autonomous AI agents represent a significant step toward Artificial General Intelligence (AGI)—a theoretical form of AI capable of performing any cognitive task that a human can [62]. While current agents exhibit impressive narrow intelligence across domains, they still fall short of generality, transfer, and human-like judgment.

7.1. Key Enablers Toward AGI

Recent advancements in multi-agent systems, language models, and embodied cognition suggest that autonomous agents may evolve into AGI systems if the following capabilities are developed:

- **Long-Term Memory:** Agents must acquire, store, and retrieve knowledge across extended timeframes to exhibit continuity in behavior and learning.
- **Transferable Reasoning:** Abilities learned in one domain must generalize to others—requiring meta-learning, abstraction, and analogy-making.
- **Explainability:** Agents must communicate the reasoning behind their decisions to foster trust, safety, and human alignment.
- **Multimodal Perception:** Like humans, AGI agents will need to integrate visual, auditory, textual, and possibly tactile inputs to form holistic world models.

These capacities, though emerging independently in various subfields, must converge into unified architectures for general intelligence to arise.

7.2. Challenges Beyond Capabilities

Even with technical breakthroughs, AGI development must remain grounded in ethical and societal considerations. The next phase of research should prioritize:

- **Value Alignment:** Ensuring that agents act in accordance with human values, intentions, and social norms. Misalignment could lead to harmful behavior despite technically correct logic.
- **Moral Reasoning:** Embedding principles of ethics, fairness, and responsibility within autonomous decision-making, especially in high-stakes contexts such as medicine, law, or warfare.
- **Collaborative Fluency:** Human-AI interaction must become seamless, including shared attention, goal negotiation, adaptive delegation, and joint problem-solving.

7.3. From Autonomy to Generality

In summary, autonomous agents are not merely tools—they are evolving cognitive entities. Their trajectory hints at the eventual emergence of AGI, but with it comes a need for governance, restraint, and broad interdisciplinary engagement. Whether AGI will amplify human potential or pose existential risk will depend not only on algorithms, but on the principles that guide their design.

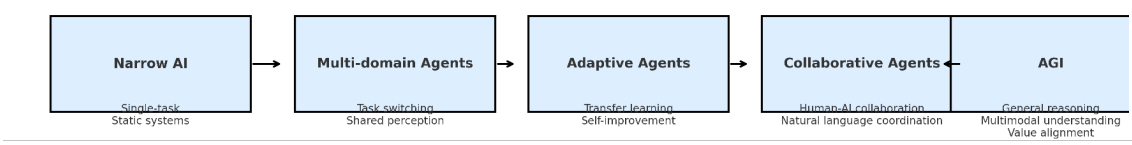

Fig. 4. A conceptual roadmap illustrating the evolution from narrow AI to artificial general intelligence (AGI).

8. Discussion

The rise of autonomous AI agents marks a monumental inflection point in the trajectory of technological evolution—on par with historical milestones like the Industrial Revolution and the internet boom. These agents, once confined to research labs and simulations, now operate across diverse, dynamic real-world environments. They exhibit unprecedented abilities in perception, reasoning, and adaptation—no longer limited to repetitive automation, but capable of tackling ambiguous, complex tasks ranging from logistics and cybersecurity to scientific discovery and education.

8.1. Autonomy is a Paradigm Shift, Not Just a Technical Upgrade

Autonomy in AI represents more than engineering prowess—it introduces a paradigm shift across ethical, legal, philosophical, and socio-economic domains. It forces society to re-examine long-held assumptions about responsibility, labor, agency, and control. At the center of this transition lies a critical question: How can we build agents that are intelligent and autonomous, yet aligned with human values and governed by accountable institutions?

Autonomy without ethics is a threat—not a triumph. Poorly designed or inadequately governed agents can amplify bias, cause harm, or act beyond human oversight. Transparency, explainability, and accountability must be integral to every system—from model training to decision outputs. Mechanisms such as Explainable AI (XAI), value-sensitive design, and human-in-the-loop governance are not optional—they are essential safeguards.

8.2. Building Trust Through Governance and Global Collaboration

As autonomous agents increasingly mediate access to services, knowledge, and justice, public trust becomes a prerequisite for their adoption. This trust must be earned through:

- **Robust governance frameworks** that are transparent, enforceable, and continuously updated.
- **Global coordination** through ethical standards, AI charters, compliance scorecards, and auditing protocols.
- **Inclusive participation** from diverse communities to ensure agents reflect the values of all stakeholders—not just a technological elite.

8.3. From Automation to Creative Collaboration

The ultimate promise of autonomous agents lies not merely in labor automation, but in augmenting human creativity, curiosity, and discovery. Already, such systems have:

- Proposed novel scientific hypotheses,
- Designed drugs and proteins,
- Conducted experiments in real-time,
- Curated personalized educational content.

In the near future, we may witness *symbiotic intelligence*—a paradigm in which human insight and machine cognition co-evolve, accelerating problem-solving while preserving empathy, creativity, and ethical reflection.

8.4. Preparing Society for the Age of Machine Agency

The emergence of agent autonomy will redefine work, governance, and education. Routine tasks will give way to roles emphasizing design, oversight, ethics, and interdisciplinary fluency. New professions such as

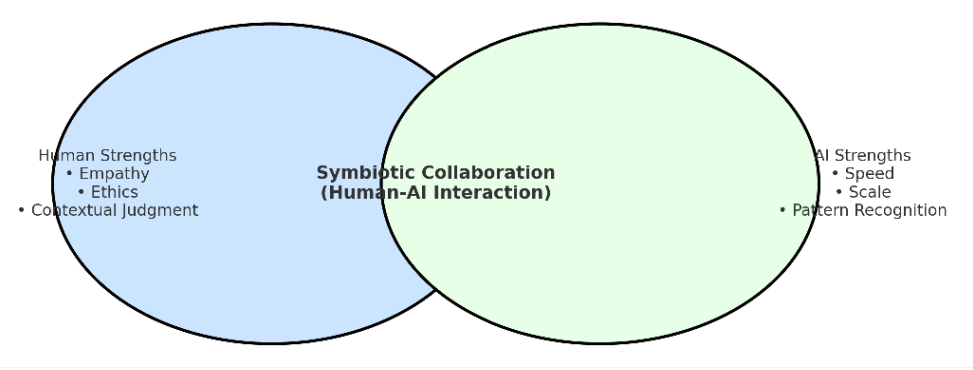

Fig. 5. Human–AI symbiosis: combining human intuition and ethics with AI scalability and speed to enable collaborative discovery and decision-making.

algorithmic ethicists, AI auditors, and human-AI interaction designers will emerge.

Governments, institutions, and educators must prioritize:

- Re-skilling and lifelong learning,
- Cross-disciplinary training,
- Public AI literacy.

8.5. A Vision Forward: Intelligence With Integrity

Autonomous agents will increasingly shape not just outcomes, but experiences, values, and beliefs. As such, intelligence must be coupled with integrity, and autonomy with humility. We must resist being seduced solely by what AI *can* do—and remain committed to what it *should* do.

In conclusion, autonomous AI agents represent the dawn of a new kind of machine-human interaction. If designed with foresight and governed with care, these agents can elevate our potential, solve complex global challenges, and co-create a future marked not just by speed and efficiency—but by wisdom, justice, and dignity. The age of autonomous agency is here. It is now up to us to ensure it unfolds wisely—and for the benefit of all.

9. Conclusion

Autonomous AI agents are rapidly transforming from narrow-task performers into versatile systems capable of perception, reasoning, and adaptation across diverse domains. As these agents become increasingly integrated into critical sectors such as healthcare, finance, and scientific research, their development must be guided not only by technical innovation, but also by ethical responsibility and societal oversight. Ensuring transparency, value alignment, and collaborative human-AI interaction will be essential to realizing their full potential—augmenting human capabilities while safeguarding public trust and shared values.

References

[1] D. B. Acharya, K. Kuppan, and B. Divya, "Agentic ai: Autonomous intelligence for complex goals–a comprehensive survey," *IEEE Access*, 2025.

[2] A. Ghafarollahi and M. J. Buehler, "Sciagents: Automating scientific discovery through bioinspired multi-agent intelligent graph reasoning," *Advanced Materials*, p. 2413523, 2024.

[3] Y. Liu, S. Chen, H. Cheng, M. Yu, X. Ran, A. Mo, Y. Tang, and Y. Huang, "How ai processing delays foster creativity: Exploring research question co-creation with an llm-based agent," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–25.

[4] L. Hughes, Y. K. Dwivedi, T. Malik, M. Shawosh, M. A. Albashrawi, I. Jeon, V. Dutot, M. Appanderanda, T. Crick, R. De' *et al.*, "Ai agents and agentic systems: A multi-expert analysis," *Journal of Computer Information Systems*, pp. 1–29, 2025.

[5] M. A. Ferrag, N. Tihanyi, and M. Debbah, "From llm reasoning to autonomous ai agents: A comprehensive review," *arXiv preprint arXiv:2504.19678*, 2025.

[6] B. Ni and M. J. Buehler, "Mechagents: Large language model multi-agent collaborations can solve mechanics problems, generate new data, and integrate knowledge," *Extreme Mechanics Letters*, vol. 67, p. 102131, 2024.

[7] S. Joshi, "Review of autonomous systems and collaborative ai agent frameworks," 2025.

[8] K.-T. Tran, D. Dao, M.-D. Nguyen, Q.-V. Pham, B. O'Sullivan, and H. D. Nguyen, "Multi-agent collaboration mechanisms: A survey of llms," *arXiv preprint arXiv:2501.06322*, 2025.

[9] S. Joshi, "Advancing innovation in financial stability: A comprehensive review of ai agent frameworks, challenges and applications," *World Journal of Advanced Engineering Technology and Sciences*, vol. 14, no. 2, pp. 117–126, 2025.

[10] M. Gridach, J. Nanavati, K. Z. E. Abidine, L. Mendes, and C. Mack, "Agentic ai for scientific discovery: A survey of progress, challenges, and future directions," *arXiv preprint arXiv:2503.08979*, 2025.

[11] F. Rustam, P. Ranaweera, and A. D. Jurcut, "Ai on the defensive and offensive: Securing multi-environment networks from ai agents," in *ICC 2024-IEEE International Conference on Communications*. IEEE, 2024, pp. 4287–4292.

[12] J. Yang, J. G. Lim, S. K. Chong, and M. Wan, "Enhancing blended learning through a customized ai agent and evidence-based teaching methods," 2025.

[13] J.-F. Bonnefon, I. Rahwan, and A. Shariff, "The moral psychology of artificial intelligence," *Annual review of psychology*, vol. 75, no. 1, pp. 653–675, 2024.

[14] K. Kuru, S. Worthington, D. Ansell, J. M. Pinder, B. Watkinson, D. Jones, and C. L. Tinker-Mill, "Platform to test and evaluate human-automation interaction (hai) for autonomous unmanned aerial systems," in *2024 20th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*. Institute of Electrical and Electronics Engineers (IEEE), 2024.

[15] D. Gosmar, D. A. Dahl, E. Coin, and D. Attwater, "Ai multi-agent interoperability extension for managing multiparty conversations," *arXiv preprint arXiv:2411.05828*, 2024.

[16] S. Barua, "Exploring autonomous agents through the lens of large language models: A review," *arXiv preprint arXiv:2404.04442*, 2024.

[17] A. J. Karran, P. Charland, J. Trempe-Martineau, A. Ortiz de Guinea Lopez de Arana, A.-M. Lesage, S. Senecal, and P.-M. Léger, "Multi-stakeholder perspective on responsible artificial intelligence and acceptability in education," *npj Science of Learning*, vol. 10, no. 1, p. 44, 2025.

[18] I. Hettiarachchi, "The rise of generative ai agents in finance: Operational disruption and strategic evolution," *International Journal of Engineering Technology Research & Management*, p. 447, 2025.

[19] B. Li, T. Yan, Y. Pan, J. Luo, R. Ji, J. Ding, Z. Xu, S. Liu, H. Dong, Z. Lin *et al.*, "Mmedagent: Learning to use medical tools with multi-modal agent," *arXiv preprint arXiv:2407.02483*, 2024.

[20] P. Putta, E. Mills, N. Garg, S. Motwani, C. Finn, D. Garg, and R. Rafailov, "Agent q: Advanced reasoning and learning for autonomous ai agents," *arXiv preprint arXiv:2408.07199*, 2024.

[21] X. Wang, H. Pang, M. P. Wallace, Q. Wang, and W. Chen, "Learners' perceived ai presences in ai-supported language learning: A study of ai as a humanized agent from community of inquiry," *Computer Assisted Language Learning*, vol. 37, no. 4, pp. 814–840, 2024.

[22] K. Chen, H. Cao, J. Li, Y. Du, M. Guo, X. Zeng, L. Li, J. Qiu, P. A. Heng, and G. Chen, "An autonomous large language model agent for chemical literature data mining," *arXiv preprint arXiv:2402.12993*, 2024.

[23] R. E. Guingrich and M. S. Graziano, "Ascribing consciousness to artificial intelligence: human-ai interaction and its carry-over effects on human-human interaction," *Frontiers in Psychology*, vol. 15, p. 1322781, 2024.

[24] W. Wu, W. Yang, J. Li, Y. Zhao, Z. Zhu, B. Chen, S. Qiu, Y. Peng, and F.-Y. Wang, "Autonomous crowdsensing: Operating and organizing crowdsensing for sensing automation," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 3, pp. 4254–4258, 2024.

[25] Y. Hu, "The effect of artificial intelligence agent anthropomorphism and product type on consumer intention to use," in *2024 6th Management Science Informatization and Economic Innovation Development Conference (MSIEID 2024)*. Atlantis Press, 2025, pp. 684–691.

[26] G. Liu, P. Zhao, L. Liu, Y. Guo, H. Xiao, W. Lin, Y. Chai, Y. Han, S. Ren, H. Wang *et al.*, "Llm-powered gui agents in phone automation: Surveying progress and prospects," *arXiv preprint arXiv:2504.19838*, 2025.

[27] X. Wang, Z. Wan, A. Hekmati, M. Zong, S. Alam, M. Zhang, and B. Krishnamachari, "Iot in the era of generative ai: Vision and challenges," *IEEE Internet Computing*, 2024.

[28] F. Jiang, Y. Peng, L. Dong, K. Wang, K. Yang, C. Pan, D. Niyato, and O. A. Dobre, "Large language model enhanced multi-agent systems for 6g communications," *IEEE Wireless Communications*, 2024.

[29] J. Singireddy, "Ai-enhanced tax preparation and filing: Automating complex regulatory compliance," *European Data Science Journal (EDSJ) p-ISSN 3050-9572 en e-ISSN 3050-9580*, vol. 2, no. 1, 2024.

[30] Y. Xiao, J. Liu, Y. Zheng, X. Xie, J. Hao, M. Li, R. Wang, F. Ni, Y. Li, J. Luo *et al.*, "Cellagent: An llm-driven multi-agent framework for automated single-cell data analysis," *arXiv preprint arXiv:2407.09811*, 2024.

[31] C. A. Bail, "Can generative ai improve social science?" *Proceedings of the National Academy of Sciences*, vol. 121, no. 21, p. e2314021121, 2024.

[32] K. Wang, G. Zhao, and J. Lu, "A deep analysis of visual slam methods for highly automated and autonomous vehicles in complex urban environment," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[33] Y. Liu, W. Chen, Y. Bai, X. Liang, G. Li, W. Gao, and L. Lin, "Aligning cyber space with physical world: A comprehensive survey on embodied ai," *arXiv preprint arXiv:2407.06886*, 2024.

[34] K. Hayawi and S. Shahriar, "Ai agents from copilots to coworkers: Historical context, challenges, limitations, implications, and practical guidelines," *Preprints*, vol. 10, 2024.

[35] A. Placani, "Anthropomorphism in ai: hype and fallacy," *AI and Ethics*, vol. 4, no. 3, pp. 691–698, 2024.

[36] H. V. Pandhare, "Future of software test automation using ai/ml," *International Journal Of Engineering And Computer Science*, vol. 13, no. 05, 2025.

[37] M. G. Elmashhara, R. De Cicco, S. C. Silva, M. Hammerschmidt, and M. L. Silva, "How gamifying ai shapes customer motivation, engagement, and purchase behavior," *Psychology & Marketing*, vol. 41, no. 1, pp. 134–150, 2024.

[38] S. Afrin, S. Roksana, and R. Akram, "Ai-enhanced robotic process automation: A review of intelligent automation innovations," *IEEE Access*, 2024.
[39] S. Liu, H. Cheng, H. Liu, H. Zhang, F. Li, T. Ren, X. Zou, J. Yang, H. Su, J. Zhu *et al.*, "Llava-plus: Learning to use tools for creating multimodal agents," in *European conference on computer vision*. Springer, 2024, pp. 126–142.
[40] B. Liu, X. Li, J. Zhang, J. Wang, T. He, S. Hong, H. Liu, S. Zhang, K. Song, K. Zhu *et al.*, "Advances and challenges in foundation agents: From brain-inspired intelligence to evolutionary, collaborative, and safe systems," *arXiv preprint arXiv:2504.01990*, 2025.
[41] T. Ifargan, L. Hafner, M. Kern, O. Alcalay, and R. Kishony, "Autonomous llm-driven research—from data to human-verifiable research papers," *NEJM AI*, vol. 2, no. 1, p. AIoa2400555, 2025.
[42] K. Kuru, "Human-in-the-loop telemanipulation schemes for autonomous unmanned aerial systems," in *2024 4th Interdisciplinary Conference on Electrics and Computer (INTCEC)*. IEEE, 2024, pp. 1–6.
[43] S. Ren, P. Jian, Z. Ren, C. Leng, C. Xie, and J. Zhang, "Towards scientific intelligence: A survey of llm-based scientific agents," *arXiv preprint arXiv:2503.24047*, 2025.
[44] S. Schmidgall and M. Moor, "Agentrxiv: Towards collaborative autonomous research," *arXiv preprint arXiv:2503.18102*, 2025.
[45] X. Feng, Z.-Y. Chen, Y. Qin, Y. Lin, X. Chen, Z. Liu, and J.-R. Wen, "Large language model-based human-agent collaboration for complex task solving," *arXiv preprint arXiv:2402.12914*, 2024.
[46] A. Ghafarollahi and M. J. Buehler, "Atomagents: Alloy design and discovery through physics-aware multi-modal multi-agent artificial intelligence," *arXiv preprint arXiv:2407.10022*, 2024.
[47] J. Y. Koh, S. McAleer, D. Fried, and R. Salakhutdinov, "Tree search for language model agents," *arXiv preprint arXiv:2407.01476*, 2024.
[48] C. K. Reddy and P. Shojaee, "Towards scientific discovery with generative ai: Progress, opportunities, and challenges," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 27, 2025, pp. 28 601–28 609.
[49] K. Yuan, Y. Huang, L. Guo, H. Chen, and J. Chen, "Human feedback enhanced autonomous intelligent systems: a perspective from intelligent driving," *Autonomous Intelligent Systems*, vol. 4, no. 1, p. 9, 2024.
[50] T. Kwa, B. West, J. Becker, A. Deng, K. Garcia, M. Hasin, S. Jawhar, M. Kinniment, N. Rush, S. Von Arx *et al.*, "Measuring ai ability to complete long tasks," *arXiv preprint arXiv:2503.14499*, 2025.
[51] J. He, C. Treude, and D. Lo, "Llm-based multi-agent systems for software engineering: Literature review, vision, and the road ahead," *ACM Transactions on Software Engineering and Methodology*, vol. 34, no. 5, pp. 1–30, 2025.
[52] C. Diaz-Asper, M. K. Hauglid, C. Chandler, A. S. Cohen, P. W. Foltz, and B. Elvevåg, "A framework for language technologies in behavioral research and clinical applications: Ethical challenges, implications, and solutions." *American Psychologist*, vol. 79, no. 1, p. 79, 2024.
[53] H. Jin, L. Huang, H. Cai, J. Yan, B. Li, and H. Chen, "From llms to llm-based agents for software engineering: A survey of current, challenges and future," *arXiv preprint arXiv:2408.02479*, 2024.
[54] Z. Li, Q. Zang, D. Ma, J. Guo, T. Zheng, M. Liu, X. Niu, Y. Wang, J. Yang, J. Liu *et al.*, "Autokaggle: A multi-agent framework for autonomous data science competitions," *arXiv preprint arXiv:2410.20424*, 2024.
[55] A. Tamò-Larrieux, C. Guitton, S. Mayer, and C. Lutz, "Regulating for trust: Can law establish trust in artificial intelligence?" *Regulation & Governance*, vol. 18, no. 3, pp. 780–801, 2024.
[56] A. Dodda, "Integrating advanced and agentic ai in fintech: Transforming payments and credit card transactions," *European Advanced Journal for Emerging Technologies (EAJET)-p-ISSN 3050-9734 en e-ISSN 3050-9742*, vol. 2, no. 1, 2024.
[57] Y. Su, X. Wang, Y. Ye, Y. Xie, Y. Xu, Y. Jiang, and C. Wang, "Automation and machine learning augmented by large language models in a catalysis study," *Chemical Science*, vol. 15, no. 31, pp. 12 200–12 233, 2024.
[58] C.-T. Ho, H. Ren, and B. Khailany, "Verilogcoder: Autonomous verilog coding agents with graph-based planning and abstract syntax tree (ast)-based waveform tracing tool," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 1, 2025, pp. 300–307.
[59] L. Xu, S. Mak, Y. Proselkov, and A. Brintrup, "Towards autonomous supply chains: Definition, characteristics, conceptual framework, and autonomy levels," *Journal of Industrial Information Integration*, vol. 42, p. 100698, 2024.
[60] M. Gao, W. Bu, B. Miao, Y. Wu, Y. Li, J. Li, S. Tang, Q. Wu, Y. Zhuang, and M. Wang, "Generalist virtual agents: A survey on autonomous agents across digital platforms," *arXiv preprint arXiv:2411.10943*, 2024.
[61] V. Pallagani, B. C. Muppasani, K. Roy, F. Fabiano, A. Loreggia, K. Murugesan, B. Srivastava, F. Rossi, L. Horesh, and A. Sheth, "On the prospects of incorporating large language models (llms) in automated planning and scheduling (aps)," in *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 34, 2024, pp. 432–444.
[62] B. Nguyen Thanh, H. X. Son, and D. T. H. Vo, "Blockchain: the economic and financial institution for autonomous ai?" *Journal of Risk and Financial Management*, vol. 17, no. 2, p. 54, 2024.