

A scientometric analysis of AI agent research: trends, applications, and future directions

Hasti Bagherzadi^{a*}

^a*Schulich School of Business, York University, Toronto, Canada*

CHRONICLE

ABSTRACT

Article history:

Received: November 3, 2025
 Received in revised format: November 3, 2025
 Accepted: November 31, 2025
 Available online:
 November 31, 2025

Keywords:

AI agents
Scientometric analysis
Large language models
Agentic AI
Autonomous systems
Human-AI interaction

The swift development of artificial intelligence (AI) has shifted to a new paradigm, i.e., AI agents—-independent entities that can sense, think and do different things in ever-changing surroundings. This paper, through a scientometric analysis, looks at the growing territory of AI agents' research by sifting through data related to a wide-ranging dataset of academic articles. It maps out the key research directions, the most popular areas for application, the prominent methodological approaches, and the nascent difficulties. The results reveal that there has been a considerable increase in the research related to AI agents, which is mainly due to the progress made in the field of large language models (LLMs), multimodal AI, and agentic frameworks. The most important application fields are healthcare, education, manufacturing, finance, and smart cities. The research also points to the limitations in terms of ethics, security, and operations that would need to be worked through if AI agents are to be deployed in a responsible manner. This study not only presents an organized picture of the current situation, but also indicates new areas for researchers to explore.

© 2025 by the authors; licensee Growing Science, Canada.

1. Introduction

The development of AI agents from theoretical frameworks to practical systems is a major event in the timeline of artificial intelligence research. AI agents are those defined as entities that perceive their surroundings and act to achieve certain objectives. They have moved from simple, rule-based, and reactive models to intelligent, sophisticated, and context-aware systems. This change has been pushed forward through innovations in machine learning, natural language processing, and robotics, which all work together to allow agents to extract meaning from the data, engage in learning through interactions, and take decisions on their own. The early applications were slow-paced by rigid coding and a little bit of flexibility, often working only in very specific areas. The rise of artistic AI, particularly large language models (LLMs), has nearly doubled the power of AI agents. These models give the agents the ability to conduct a challenging conversation, think about complex situations, and constantly change their behavior in accordance with the changing inputs (Sapkota et al., 2026; Ren et al., 2025). So, AI agents are not now only in laboratories but also in the real world, where they can be found in various industries and used in a variety of applications. They are helping in diagnosing and monitoring patients in the healthcare sector; they are making the learning process more engaging in the education sector; they are increasing the efficiency of production in industrial automation; and they are facilitating the development of smart infrastructure and proper resource allocation in the urban management sector. The combination of generative intelligence and agent-based autonomy marks the beginning of a new age of intelligent systems that will significantly affect society.

This scientometric review plans to map the intellectual structure of AI agent studies by analyzing a curated corpus of recent publications. Unlike traditional literature reviews, this study employs a quantitative and qualitative analysis of metadata—including authorship, affiliations, keywords, and abstracts—to identify patterns, collaborations, and knowledge gaps. The primary research questions are:

* Corresponding author.

E-mail address: Hastib@yorku.ca (H. Bagherzadi)

1. What are the dominant themes and subfields within AI agent research?
2. Which application domains are most frequently addressed?
3. What methodological approaches are prevalent?
4. What are the emerging challenges and future research directions?

The dataset consists of 50 articles, which were obtained through Scopus, all of which have the term "AI agent" in their titles, and they have been mainly published in 2025 and 2026. This selection represents the most advanced research on AI agents and provides a current view of the development of the field.

2. Methodology

This scrutiny revolves around a meticulously arranged dataset that has been derived from Scopus, which grants access to the bibliographic records of the publications within the period of 2024-2026. The dataset entry has information about authors, titles, abstracts, keywords, affiliations, and citation statistics. The next bibliometric methods were utilized:

- **Thematic Analysis:** Identification of recurring topics through author keywords and abstract mining.
- **Citation Analysis:** Examination of cited references and citation counts to gauge influence.
- **Collaboration Mapping:** Analysis of co-authorship networks and institutional partnerships.
- **Content Categorization:** Classification of articles by application domain, methodology, and agent type.

We did a manual search because the dataset was quite small, allowing us to read into context and spot only minor shifts as per tendencies. Dropout versions are effective in selecting and annotating features (classes) in that if a dimension of problem attribute is not pertained supposedly to feature (class) selection, the feature (class) gets removed by the attribute reduction algorithm (BBAs).

3. Thematic Clusters in AI Agent Research

The literature discloses various coherent thematic clusters, each representing a distinct research focus.

3.1. Foundation and Theory of AI Agents

The rapid growth of the field along with its growing complexity has resulted in a lot of scholarly work in the area of conceptualization and classification of AI agents. One of the most important contributions, Sapkota et al. (2026), provide a very important division between AI Agents and Agentic AI that has led to a taxonomy which is still very much shaping the discourse. Their framework classifies AI Agents as system components enhanced by LLMs and LIMs and used mostly for task-specific automation. On the other hand, Agentic AI heralds a paradigm shift by pointing to multi-agent cooperation, fluid task splitting, and persistent memory use, thus supporting long-term autonomy and adaptability. The distinction has been underfundamentally elicited and become the basis of discourse, thus offering a way for researchers to view the case of the emerging architectures and capabilities. Centering on the given base, Ren et al. (2025) research deeper into the generative AI's paradigm shift concerning the semantic comprehension and the autonomous intra-logic. They come up with the terms of LLM-Agents and MLLM-Agents, who incorporate the language and the multimodal competences, respectively. It is these agents that are being ever more widely used in the smart manufacturing sector, where domain-oriented customization and instantaneous decision-making are of utmost significance. Together, these studies reveal the change of focus from static automation to dynamic, context-aware intelligence thus highlighting the new chapter in the progress of agent-based systems.

3.2. Human-AI Interaction and Social Aspects

A growing body of research from various fields is revealing how humans perceive, trust, and interact with AI agents, thus exposing the psychological, ethical and sociocultural aspects of these relationships. Lei et al. (2025) present strong support for the idea that the perception of an AI agent's role—either as a “servant” or a “partner”—has a major impact on consumer behavior. Their research indicates that agents with a servant-like personality can create an expectation of moral disengagement and, as a result, make users less reluctant to justify their immoral actions. On the other hand, partner-like agents give rise to a feeling of shared responsibility and ethical restraint which, in turn, lessens such behaviors. This points to the importance of agent framing and relationship cues in the process of human decision-making and moral accountability.

Zhang et al. (2025) investigate the tricky area of AI empathy through the lens of the user's cognitive and affective dimensions in empathy, how they impact the user's interaction with the AI. The study results show that cognitive empathy—agent understanding of the user intent—improves perceived novelty and intellectual involvement. At the same time, affective empathy, which shows the emotional connection, has an impact on the perceived warmth and eeriness, thus affecting the behavior of the users in either the approach or the avoidance direction. These findings are especially important for the area of mental health where AI agents need to be empathetic enough to be perceived as emotionally-loaded but also user-friendly, and for the field of conversational AI where the lack of subtlety in emotional cues could lead to loss or decline of user trust and retention.

Borau (2025) proposes a very important ethical viewpoint by considering the consequences of female-gendered AI agents from an ethical perspective. She contends that such designs would not only reinforce gender stereotypes but also facilitate manipulation and maintain unequal distribution of power. Her studies demand more transparency, inclusive design practices, and managerial ethics over the development of agents, particularly since AI agents will soon be universally interacted with on a daily basis. All in all, these studies point to the necessity of human-centered design where psychological realism, ethical integrity, and social sensitivity drive the making of AI agents that not only possess intelligence but also are responsible and esteem human values.

3.3. Architectural and Technical Advances

The recent studies have gradually shifted the focus towards the technical architecture and operational potentials of AI agents, thus proving their increasing sophistication in various fields. Ding et al. (2026) present the CCM-FCC framework as a human-robot collaboration enhancement model centered on cognition. Their method, which blends the Chain-of-Thought prompting with functional cluster collaboration, gives AI agents the capacity to generalize and to be in tune with the surroundings, thus being proactive and context-aware in interaction—this is a major step forward in interaction capabilities. When talking about digital twins, Wang et al. (2025) suggest a virtual in-situ calibration system where AI agents are the main characters. The automation of sensor calibration through the use of the Brick schema ontology enhances the accuracy and reliability of digital twin models. Yoon et al. (2025) corroborate this notion by putting forward the ontology-enabled AI agent framework for building operations and maintenance. These agents exercise reasoning at a high level and can carry out tasks independently, which is an indication of the significant contribution of AI to the management of intelligent infrastructure. Another area that Huang et al. (2025) discuss is recommendation systems, and they are the creators of InteRec-Agent, a hybrid model that mixes LLMs with traditional recommender algorithms. The combination allows for interactive, conversational engagement as well as domain-specific precision. Each of the above-mentioned studies individually, yet collectively, emphasize the multiple capabilities and technical aspects of AI agents in the real-world scenario.

3.4. Security and Trustworthiness

AI agents are increasingly being given more freedom and power to make decisions, which has led to the concerns regarding security and trust to become even more pressing. A study conducted by Deng et al. (2025) gives an overall view of the security weaknesses in the agent systems and at the same time bringing up the critical risks related to several aspects like user inputs in multiple steps, internal executions, operational environments, and working with untrusted external parties. One of the most important conclusions they reached is that there must be powerful protective measures in place to avoid exploitation and to keep the system sound, particularly because the agents are operating in unpredictable and ever-changing situations. Wang et al. (2025) discuss the MCP or Model Context Protocol, which is a basic element in a lot of agent architectures. The study points out the security insecurities in the stages of input handling, decision-making, and tool invocation, especially to the threat of tool description injection, which can make the agent act in a certain way by giving false metadata. In order to close these gaps, the authors put forward a set of measures which are primarily aimed at enhancing the protocol's ability to withstand attacks and at protecting the agent's functioning. Thurzo (2025) proposes a new paradigm of ethical assurance with the "Ethical Firewall" concept - an architecture that blends formal verification with cryptographic immortality. With this, AI agents' decisions are kept to be in accordance with human values, thereby providing proofs of ethics and being able to explain things in areas that are sensitive, like health, and education. The mentioned studies combined, point to the great necessity for secure, transparent, and ethically grounded AI agents especially in high-stakes contexts where trust is the overriding factor.

4. Application Domains

AI agent research is highly applied, with studies spanning numerous domains.

4.1. Healthcare and Mental Health

The healthcare sector is witnessing a rising trend in the adoption of AI agents, which again is a great advantage for apps due to their ability to handle complicated data and to use natural language in the interaction. In a recent work by Obadinma et al. (2025), FAIR is presented as a main actor in this field whose role is to support youth mental health services with the

aid of a conversational AI agent that is especially designed for this purpose. The development consists of the combination of several transformers that are powerful FAIR helps professionals in crises by pinpointing problems and providing resource allocation in real time. The system claims to provide a very reliable output as it matches expert opinions and shows a high recall. On the other hand, Zhao et al. (2025) developed a clinical setting for their research by conducting a randomized controlled trial to assess the LLM-based AI agent for the treatment of depression and anxiety. The four weeks of dialog-based interaction with the participants had resulted in the significant lowering of symptoms, hence the conclusion that the availability of conversational agents can provide a larger, easier-to-access mental health support at a lower cost. The main point of the study is the AI agents' therapeutic potential to be more pronounced in areas or cases where the number of resources available is limited. The field of oncology has also seen the light with the introduction of Yang et al. (2025) who have developed a specialized conversational agent primarily targeting the TGF- β pathway alterations in colorectal cancer termed AI-HOPE-TGFbeta. The system permits doctors and scientists to conduct a natural language-based query on both clinical and genomic data which thus permits the analysis by integration and the acceleration of precision oncology efforts. All together the spreading of such innovations signifies a trend of rising domain-specific, intelligent agents that all the more so will support and handle clinical decision-making and patient care.

4.2. Education and Training

AI agents are progressively being merged into the educational setup for the improvement of both assessment and learning. Yang et al. (2025) present a Creative Project Approach where AI agents are coupled with robots in early childhood education. The authors provide a comprehensive pedagogical framework around the interaction between the child and the robot in a guided AI integration to promote creativity, exploration, and collaborative learning. This technique not only nurtures cognitive growth but also technologizes very young pupils to basic concepts of humankind and machine partnership. On the contrary, Yan et al. (2025) in the realm of higher education, investigate the impact of generative AI agents on the students' grasp of visual learning analytics. Their research indicates that proactive agents—who play an active role in guiding and responding to learners—are a crucial factor in comprehension that is considerably higher as compared to the use of passive agents or conventional scaffolding methods. These conclusions imply that interactive AI agents can act as powerful facilitators in data-driven learning environments, making it easier for students to process complex visualizations and gain valuable insights. Tyndall et al. (2025), on the other hand, investigate the use of AI agents in the examination process from creating the questions to answering them at the undergraduate level. They show that the retrieval-augmented generation (RAG) technique not only improves the accuracy but also coherence of answering questions by providing the agents with the ability to generate contextually relevant responses based on the external knowledge. This study indicates a future where AI instructors not only assist learners but also engage in the grading process, thus, providing educational tools that are colossal, adaptive, and intelligent.

4.3. Manufacturing and Robotics

AI agents are turning out to be more and more important in the industrial settings where the traits of adaptability, preciseness, and automation are valued the most. Ding et al. (2026) have discussed the proactive human-robot teamwork within the difficult tasks of disassembly and assembly. The authors have come up with the Cognition-Centered Model with Functional Cluster Collaboration (CCM-FCC) framework that allows the robots to automatically adjust their reactions according to operator states and the level of difficulty of tasks. Combining semantic reasoning and modular task coordination, this architecture not only increases the real-time responsiveness but also the operational efficiency, thus making it fit for the manufacturing sector where there is high variance. On the other hand, Fan et al. (2025) present AutoMEX, which is an AI agent framework specifically designed for the material extrusion additive manufacturing. AutoMEX uses a knowledge graph along with large language models (LLMs) to automate vital production pipeline stages like material selection, parameter tuning, and printing execution. This smart orchestration leads to a significant cut in manual intervention, a decrease in errors, and an increase in output consistency. The system's capability of understanding domain-specific knowledge and adapting to new materials or design requirements is a clear example of the AI-driven customization trend in industrial workflows. Innovations such as these are markers of a general move to intelligent automation where AI agents perform not only the given tasks but also interact, think, and optimize the processes simultaneously. This progress is altering the industrial scene, turning factories into more intelligent ones and production systems into more robust ones.

4.4. Smart Cities and Urban Systems

AI agents are continuously being brought into the cities to help deal with the problems of energy consumption, transportation, and infrastructure. Choi and Yoon (2025) suggest the Intelligent Urban Digital Twin (I-UDT) which is an AI agent-based system for building energy modeling. The I-UDT, through the integration of GPT-based analysis, optimizes energy consumption patterns and supports carbon-neutral strategies providing a scalable solution for sustainability in urban development. This is one instance of using generative AI that can contribute to the development and improvement of predictive modeling and decision-making in challenging and complicated urban systems. Jiao and Chang (2025) utilize AI agents to analyze the sentiments and issues that have been expressed in user reviews of electric vehicle (EV) charging stations, and as a result, discover spatial patterns in sentiment and operational problems. Their approach leads to the urban planners being

able to recognize where services are lacking, where there is a disparity among regions, and where there are bottlenecks in infrastructure; thus, informing urban planning and service optimization. The application of AI for sentiment mapping and geospatial analysis is indicative of the capability of agent-based systems in converting public feedback into insights that can be acted upon. Going even further, de Silva et al. (2025) come up with a hyper-automation framework based on generative AI agents for sustainable smart cities. Their solution makes use of the streams of real-time data being processed to support decision-making in areas such as traffic management, public safety, and resource allocation. Together these advances show how AI agents are becoming key players in urban intelligence, allowing cities to be more effective, more responsive, and more sustainable in their operations.

4.5. Finance and Business

The financial industry is slowly but surely being overtaken by AI agents in their various applications like trading, customer service, and risk management. One of the most notable instances of this is the initiative by Wang et al. (2025) who presented a specially designed AI agent for trading agricultural futures, which uses market signals interpreted through multiscale candlestick analysis and vision-language models to suggest buying or selling. This combination of methods not only improves the recognition of patterns but also the understanding of trading contexts thereby allowing more informed and adaptable strategies involving trading in the commodities market that is subject to frequent price changes. Customer service is one area where systems failures have an impact on customer behavior, and Li et al. (2025) are studying this aspect in relation to AI agents and traditional interactive voice response (IVR) systems. The results of their research indicate that when customers are assisted by a human, AI agent failures are more likely to lead to higher post-recovery purchasing than IVR failures. This implies that customers are likely to find AI agents quite interactive and responsive, and this perception becomes especially strong when the agents are backed up by a human showing empathy. Thus, the study highlights the need for having hybrid customer service models if the business aims to keep its customers. Chatterjee et al. (2025) carry out a study that helps to understand the extent to which human and AI agent services affect customer learning, immersion, and loyalty. Furthermore, this study points out the significant role of the perceived interactivity, as it is AI agents that are able to conduct engaging conversations with users that are the ones who develop deeper cognitive engagement and brand affinity. Such revelations indicate the strategic necessity of AI agents in the process of improving user experience and forming long-term customer relationships in the entire financial sector.

5. Methodological Approaches

The research employs diverse methodologies, reflecting the interdisciplinary nature of AI agent studies.

5.1. LLM Integration and Prompt Engineering

The recent progress in AI agent design has more and more incorporated large language models (LLMs) as the main reasoning engine, which has opened door to more sophisticated and autonomous capabilities in various fields. Go and Park (2025) are concerned with the issue of tool-augmented LLMs, and they present the idea of having concurrent API calls and multi-step reasoning mechanisms. As a result of their improvements, there is a significant increase in the accuracy of performing tasks by agents, this increase is from 4.4% to 9.3%. This is mainly due to the fact that the agents are able to manage large-scale workflows and communicate with external tools simultaneously. The approach taken here brings about a notable increase in both the efficiency and responsiveness of systems and applications, particularly in volatile environments where making decisions in real-time is a necessity. Peng et al. (2025) are the ones who attribute the LLMs their power in knowledge extraction, and the latter being automated, this finding was drawn from the literature on deep eutectic solvents. Their setup is one that takes technical papers and breaks them down to present the main ideas, interactions, and experiment results-the whole thing done showcases the potential of LLMs in speeding up specific areas of discovery. The literature review, which historically has been a manual process, is now made less burdensome while simultaneously allowing researchers to find insights in large amounts of data, thus revealing LLMs' application in such areas as chemistry, materials science, and medical research. The studies mentioned above, each in its own way, are shedding light on the process of LLMs transitioning from mere text generators into active reasoning engines in AI agents. The researchers have combined language comprehension with tool integration and domain adaptation, thus extending the limits of intelligent systems in both technical and scientific contexts.

5.2. Reinforcement Learning and Adaptive Systems

The recent AI agent design innovations are not only improving emotional intelligence but also making it easier to decide under uncertainty. Soman et al. (2025) come up with a new virtual therapist that is emotionally aware. The chatbot combines reinforcement learning with retrieval-augmented generation (RAG). Because of the hybrid design, the agent can learn and change its replies according to users' likes and hates, feelings, and interaction feedback. The chatbot, by giving priority to emotionally close interactions, creates a secure environment and engages the user more, making it a very strong support during the mental health care process where empathy and customization are crucial. Meanwhile, Sun et al. (2025) face the uncertainty problem in autonomous systems by using Bayesian inference and evidence accumulation theory to model belief

updates in drones. The developed framework gives the AI agents the ability to incrementally integrate probabilistic reasoning with data collection thus they will understand the environment better. The reliability of the decision-making process in fluctuating and not very clear situations like navigation, obstacle avoidance, and mission planning is greatly improved. Formalizing the manner in which agents update beliefs over time, enhances the powers of robustness and transparency—two necessary conditions for the deployment of autonomous drones in safety-critical applications. Collected together, these research works are showing the new wide area of the AI agents from the emotionally conscious chatbots that promote people's wellness to the systems that are based on probability and can work well in unclear environments. This duality of AI agents' characteristics—empathy and resilience—reflects the smart agents trend that is going on in the industry.

5.3. *Ontology and Knowledge Graphs*

Ontologies are knowledge graphs that are getting more and more important in mapping domain knowledge for AI agents, and in turn, they are offering more accurate reasoning and understanding of the context. Yoon and co-workers (2025) not only extend the Brick schema, which is a standardized ontology for building metadata but also they consider it an important step for using digital twins in the architecture and construction industry. They argue that digital twins, being the most advanced representation of buildings, allow for AI agents to comprehend even more complicated relationships at a larger scale within the environments, which could be the case with HVAC systems, lighting and occupancy patterns. Yoon et al. foresee even more enhancements in the future by embedding semantic structure into the digital twins, which would enable AI agents to carry out more accurate diagnostics, energy consumption optimization, and support of preventive maintenance—these are the main building management capabilities of the smart buildings. Zhao et al. (2025) use a likewise knowledge-driven technique in the area of traditional Chinese medicine (TCM). They couple a knowledge graph that is specific to the domain with large language models (LLMs) to help the extraction of compounds from intricate textual queries. The word hybrid system allows the agent to comprehend the subtlety of medical terminology and relational data and thus perform very accurately in identifying proper compounds and their therapeutic links. The method shows that structured knowledge can boost LLM performance in so-called specialized fields, where interpretability and precision are critical. In unison, these studies reveal the increasing interaction between ontological structures and language-driven reasoning, and at the same time, how knowledge graphs give AI agents the ability to explore complex areas with more semantic knowledge and operational dependability.

5.4. *Experimental and Empirical Validation*

Empirical validation is the backbone of portraying AI agents' effectiveness and trustworthiness in various application areas. To support their assertions, most of the recent studies make use of case studies, controlled experiments, or real-world interventions. Yu et al. (2025) explore modular AI agents through the lens of transportation surveys and disclose that these agents are the major reason for the increase in both completion rates and response quality. They imply that modularity and changing interaction methods can assist in people's active participation and correctness of the data in government services. Amazingly, Swanson et al. (2025) enable AI-driven scientific innovation and heavenly through their “Virtual Lab” a simulated research environment inhabited by AI agents up to the task of designing new SARS-CoV-2 nanobodies. These agents use computational modeling and generative design techniques to suggest molecular candidates, which subsequently are verified by laboratory experiments. Successful reaching from virtual design to experimental confirmation emphasizes the capability of AI agents in hastening biomedical discovery. All these studies provide the same conclusion, namely that it is imperative for AI research to be empirically grounded. Researchers who go beyond theoretical models and simulations can show how AI agents are capable of bringing considerable improvements in human-centered tasks as well as scientific exploration. The trend indicates that there is an increasing concern about the real-world impact where usability, scalability, and adoption are all tied to validation through deployment and experimentation.

6. **Challenges and Future Directions**

Despite rapid progress, the literature identifies several unresolved challenges.

6.1. *Ethical and Social Implications*

The ethical design and deployment of AI agents are now among the top issues of concern alongside machine systems gradually taking over the communication, decision-making, and emotional interaction of humans. The debates are focused on bias, privacy, and autonomy that are the main factors rendering the incorporation of AI into human life almost impossible unless responsible and ethical practices are followed. Borau (2025) and Lott & Hasselberger (2025) investigate human-AI interactions from moral and relational points of view, thus questioning the very nature of human-AI relationships—their meanings and ethics—considering the responsiveness and realism of such entities. They contend that the emotional reciprocity perceived is fundamentally skewed and possibly misleading because there is no real care, consciousness, or moral standing from the AI agents' side. In case users confuse agents with humans or take them as a source of emotional support (e.g., mental health, education, or companionship applications), then these concerns will be even more pronounced. The synecdoche of mutual recognition might be a curtain that hides not only the involved parties' and designers' intentions but

also the power differentials, thereby increasing the likelihoods of manipulation, dependency, or loss of human agency. Researches to come should, therefore, leadingly focus on value alignment as a measure to support the challenge of human ethical escalation, whereby AI agents act in conformity with human ethics. Another aspect that is as critical as that one is the transparency of agent capabilities and limitations and the user empowerment to individuals having the right to and being able to understand, control, and meaningfully consent to their interactions with AI. These conditions are the foundation not only for the development of intelligent AI systems but also for those that are ethically based and socially accepted.

6.2. Security and Robustness

The more AI agents are given freedom and incorporated into crucial systems, the more difficult it will be to deal with their weaknesses that come from hostile attacks, misuse of the tool, and unpredictable behaviors. Deng et al. (2025) and Wang et al. (2025) emphasize the need for secure and reliable interaction through protocols, and mechanisms for decision-making that can be verified. It is then, these protections will be able to block the malicious exploitation and unintended actions, as if the latter, the former, and the consequent ones are especially in the case of multi-agent situations or systems where external tools are accessible. Deng et al. point out the danger in multi-step communications, where the adversaries may play with the sequence of the inputs to set off dangerous actions. Wang et al. talk about the honesty of the tool use and the decision-making process, and they warn against the different types of attacks including the injection one, and when the agents get confused with the tool descriptions that are not clear. All the researchers suggest bringing in formal verification, input sanitization, and audit trails to accomplish transparency and accountability. Although the problems are recognized more and more, still, the studies of adversarial training, anomaly detection, and fail-safe mechanisms have just begun. The existing methods usually cannot be applied to various fields or cannot maintain the balance between the two different needs—robustness and flexibility. As AI agents go through and are involved in the most intricate roles ever, that of healthcare, finance, and autonomous systems—there comes a concrete demand for the interdisciplinary collaboration to create the strong structures that can identify, withstand, and recover from the disruptions caused by adversaries. This challenge is a vital one in terms of the safety, reliability, and the long-term existence of intelligent agents.

6.3. Generalization and Scalability

Artificial Intelligence agents, while often showing stunning results in controlled or simulated environments, still find it hard to generalize their behavior during open-world scenarios. In a close scrutiny of the current state of benchmarking, Kapoor et al. (2025) take a critical look at the limitations of present-day practices in benchmarking and evaluating AI performance plus they also argue that a huge percentage of evaluation frameworks weigh accuracy metrics very heavily while other vital aspects like cost, efficiency, robustness, and real-world applicability are totally overlooked. The consequences of such a limited perspective can include raising expectations to unrealistic levels and slowing down the uptake of AI agents in uncertain and fluctuating environments. The authors propose a major overhaul of the existing evaluation frameworks to make them more reflective of the real-world where the agents face difficult challenges e.g. noisy inputs, incomplete information, and evolving task requirements. Kapoor et al. (2025) recommend benchmarks that take into account the factors of resource limitations, types of failures, and long-term adaptability which will help the researchers to evaluate not only how well the agent is performing under perfect conditions but also his/her resiliency and scalability in the real world. Such change in the evaluation philosophy is absolutely necessary for the progress of the field beyond localized laboratory success to trustworthy, deployable AI agents. The embedding of the agents in various domains like healthcare, transportation, and urban infrastructure would increase their presence in the real world; hence, robust and context-sensitive benchmarking across the different domains will always be a prerequisite to ensuring the agents' reliability, safety, and contribution to the society in terms of value.

6.4. Human-AI Collaboration

The division of labor between humans and AI agents is still a matter of discussion, especially in cases where emotions, adaptability, and long-term relationships are needed. Chang and co-authors (2025) look at this problem through the perspective of support for the elderly and people with cognitive impairments, stressing the need to preserve user independence while offering significant help. Their research shows a scenario where AI agents become friends, guides, or cognitive helpers but not to the point of being intrusive or giving too much direction. One major point of their paper is the suggested changing of AI agents from "unremarkable" to "remarkable" in the light of the changing user requirements. At the very start, it is likely that agents will work more or less unnoticed but at the same time giving background support and letting the user continue with what they were doing with the least possible interference. But when the user's challenges, either cognitive or physical, become harder, the agents are going to be expected to take on more and more diverse roles, become more visible and personalized, and still respect the user's preferences and dignity. This change reflects a model of agent engagement that is needs-based, where the level of involvement changes with the user's gradual change of capabilities. Chang et al. (2025) draw attention to the need for developing the agents that are aware of the context they are operating in, emotionally sensitive, and able to gradually assume different roles. The results of their research not only underline the necessity of advanced technical skills but rather the need for implementation based on principles of considerate, user-centred design through conversation about human-AI collaboration in general.

6.5. Infrastructure and Interoperability

With the increasing use and independence of AI agents, the development of strong infrastructure that provides support for their scalability, accountability, and trustworthiness is being acknowledged more than ever before. Chan et al. (2025) point out that the foundation of such infrastructures lies in the use of attribution systems, communication protocols, and governance mechanisms which can help in creating large agent ecosystems that can be trusted easily. These foundational components make it possible for agents to carry out their transactions in a transparent manner, be blamed for their actions, and work under clearly specified moral and functional restrictions. In a slightly different perspective, Gürpınar (2025) talks about a Web 4.0 framework where non-centralized AI agents will be using blockchain technology and smart contracts for their interactions. The proposed architecture allows the agents to work independently on a network that is distributed but still has the ability to have their communication verified, keep their data intact, and establish a trust that is programmable. With the logic for transactions and verification of identity being incorporated into the infrastructure as per Gürpınar's model, it allows for the co-working of agents without the need for a central authority thus making the digitalecosystems more resilient and democratic. These different views come together to indicate that infrastructure-level innovation in the design of agents is becoming more and more important. If the agents are to be successful when moving from isolated applications to interconnected systems, they will have to be not only intelligent and adaptable but also governed by the rules, protocols, and architectures that control their behavior and interactions.

7. Conclusion

The outline of the scientific literature analysis presents an active and quickly developing area of the application of AI agents. The combination of LLMs, multimodal AI, and agentic systems has been the main reason behind the significant increase of autonomy, reasoning, and interactivity. The scope of research includes the theory and the technical architectures, as well as the wide range of application areas, such as healthcare, education, manufacturing, and smart cities. On the contrary, the area is suffering from big problems that are connected with ethics, security, generalization, and human-agent collaboration. Among the main topics of the future research there will be robust evaluation metrics, security infrastructure, and ethical guidelines developments aimed at responsible and effective deployment of AI agents. The uptake of AI agents in every day life will demand that interdisciplinary collaboration involving computer scientists, ethicists, lawyers, and social scientists is done in order to guide through the difficult landscape of both opportunities and risks. This review is grounding the primary overview which will help researchers, practitioners, and policymakers in determining the future line of AI agent technologies.

References

- Bai, S., He, H., Han, C., Yang, M., Li, Z., & Fan, W. (2025). Light Trumps Shadow? How Generative AI Agent's Language Arousal Influences Users' Interactive Willingness: Evidence From Multimodal Analysis. *IEEE Transactions on Engineering Management*, 72, 3921–3936.
- Borau, S. (2025). Deception, Discrimination, and Objectification: Ethical Issues of Female AI Agents. *Journal of Business Ethics*, 198(1), 1–19.
- Chan, A., Wei, K., Huang, S., Rajkumar, N., Perrier, E., Lazar, S., Hadfield, G. K., & Anderljung, M. (2025). Infrastructure for AI Agents. *Transactions on Machine Learning Research*, 2025-May.
- Chang, M. L., Reig, S., Lee, A., Huang, A., Simao, H., Han, N., Khanuja, N. M., Mohammad Ali, A. U., Martinez, R., Zimmerman, J., Forlizzi, J., & Steinfeld, A. (2025). Unremarkable to Remarkable AI Agent: Exploring Boundaries of Agent Intervention for Adults with and Without Cognitive Impairment. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), CSCW200.
- Chatterjee, R., George, S. R., Verma, J. S., Heggde, G., & Gadhavi, D. D. (2025). Impact of human and AI-agent services on customer learning, immersion and loyalty: the role of interactivity. *Journal of Service Theory and Practice*. Advance online publication.
- Choi, S., & Yoon, S. (2025). AI Agent-Based Intelligent Urban Digital Twin (I-UDT): Concept, Methodology, and Case Studies. *Smart Cities*, 8(1), 28.
- de Silva, D., Mills, N., Moraliyage, H., Rathnayaka, P., Wishart, S., & Jennings, A. (2025). Responsible Artificial Intelligence Hyper-Automation with Generative AI Agents for Sustainable Cities of the Future. *Smart Cities*, 8(1), 34.
- Deng, Z., Guo, Y., Han, C., Ma, W., Xiong, J., Wen, S., & Xiang, Y. (2025). AI Agents Under Threat: A Survey of Key Security Challenges and Future Pathways. *ACM Computing Surveys*, 57(7), 182.
- Ding, P., Zhang, J., Zhang, P., Li, H., & Wang, D. (2026). CCM-FCC: LLM-powered cognition-centered AI agent framework for proactive human-robot collaboration. *Robotics and Computer-Integrated Manufacturing*, 98, 103145.
- Fan, H., Huang, J., Xu, J., Zhou, Y., Fuh, J. Y. H., Lu, W. F., & Li, B. (2025). AutoMEX: Streamlining material extrusion with AI agents powered by large language models and knowledge graphs. *Materials and Design*, 251, 113644.
- Go, H., & Park, S. (2025). A study on classification based concurrent API calls and optimal model combination for tool augmented LLMs for AI agent. *Scientific Reports*, 15(1), 20579.
- Gürpınar, T. (2025). Towards web 4.0: frameworks for autonomous AI agents and decentralized enterprise coordination. *Frontiers in Blockchain*, 8, 1591907.

- Huang, X., Lian, J., Lei, Y., Yao, J., Lian, D., & Xie, X. (2025). Recommender AI Agent: Integrating Large Language Models for Interactive Recommendations. *ACM Transactions on Information Systems*, 43(4), ART96.
- Jiao, J., & Chang, A. (2025). Evaluating sentiment and spatial patterns of EV charging station user experience with AI-agents. *International Journal of Urban Sciences*. Advance online publication.
- Kapoor, S., Stroebel, B., Siegel, Z. S., Nadgir, N., & Narayanan, A. (2025). AI Agents That Matter. *Transactions on Machine Learning Research*, 2025-May.
- Lei, S., Xie, L., & Peng, J. (2025). Unethical Consumer Behavior Following Artificial Intelligence Agent Encounters: The Differential Effect of AI Agent Roles and its Boundary Conditions. *Journal of Service Research*, 28(4), 598–613.
- Li, B., Chang, Y., Liu, L., Liu, H., & Sun, J. (2025). How does AI agent (vs. IVR system) service failure impact customer purchase behavior: mediating effect of customer involvement. *Service Industries Journal*, 45(7–8), 702–720.
- Lott, M., & Hasselberger, W. (2025). With Friends Like These: Love and Friendship with AI Agents. *Topoi*. Advance online publication.
- Obadinma, S., Lachana, A., Norman, M. L., Rankin, J., Yu, J., Zhu, X., Mastropaolo, D., Pandya, D., Sultan, R., & Dolatabadi, E. (2025). The FAIRR conversational AI agent assistant for youth mental health service provision. *npj Digital Medicine*, 8(1), 243.
- Peng, X., Tew, Y. S., Zhao, K., Wang, C., Li, R., Hu, S., & Wang, X. (2025). Unlocking deep eutectic solvent knowledge through a large language model-driven framework and an interactive AI agent. *Green Chemical Engineering*, 6(4), 572–581.
- Ren, Y., Liu, Y., Ji, T., & Xu, X. (2025). AI Agents and Agentic AI—navigating a plethora of concepts for future manufacturing. *Journal of Manufacturing Systems*, 83, 126–133.
- Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2026). AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges. *Information Fusion*, 126, 103599.
- Soman, G., Judy, M. V., & Abou, A. M. (2025). Human guided empathetic AI agent for mental health support leveraging reinforcement learning-enhanced retrieval-augmented generation. *Cognitive Systems Research*, 90, 101337.
- Sun, B., You, H., Wu, J., Wang, Q., & Du, J. (2026). Belief Update Modeling for AI Agents by Human-Derived Dual-Thresholds Evidence Accumulation Process. *Journal of Computing in Civil Engineering*, 40(1), 04025116.
- Sun, C., Yang, X., Di Cicco, N., Ayassi, R., Virajit Garbhapu, V., Stavrou, P. A., Tornatore, M., Charlet, G., & Pointurier, Y. (2025). Experimental demonstration of local AI-Agents for lifecycle management and control automation of optical networks. *Journal of Optical Communications and Networking*, 17(8), C82–C92.
- Swanson, K., Wu, W., Bulaong, N. L., Pak, J. E., & Zou, J. Y. (2025). The Virtual Lab of AI agents designs new SARS-CoV-2 nanobodies. *Nature*, 646(8085), 716–723.
- Thurzo, A. (2025). Provable AI Ethics and Explainability in Medical and Educational AI Agents: Trustworthy Ethical Firewall. *Electronics*, 14(7), 1294.
- Tyndall, E., Gayheart, C., Some, A., Genz, J., Wagner, T., & Langhals, B. (2025). Impact of retrieval augmented generation and large language model complexity on undergraduate exams created and taken by AI agents. *Data and Policy*, 7, e57.
- Wang, J., Wu, J., Zhang, G., Tan, M., Chen, S., & Lin, Z. (2025). Agricultural Futures Trading Decision Using AI Agent With Multiscale Candlestick Analysis. *IEEE Transactions on Computational Social Systems*. Advance online publication.
- Wang, L., Tang, W., Tang, X., Wang, H., Xue, C., & Gan, L. (2025). Interface Layout of Different AI Agents in Human-Computer Cooperation. *International Journal of Human-Computer Interaction*. Advance online publication.
- Wang, P., Hu, Q., Mei, Q., Wang, S., Yang, Y., Guo, D., Liu, X., Hu, W., & Chen, J. (2025). Intelligent port logistics: A spatiotemporal knowledge graph and AI-agent framework for berth allocation. *Advanced Engineering Informatics*, 68, 103633.
- Wang, W., Li, S., Dong, T., Meng, Y., & Zhu, H. (2025). From Function Calls to MCPs for Securing AI Agent Systems: Architecture, Challenges and Countermeasures. *ZTE Communications*, 23(3), 27–37.
- Yan, L., Maldonado, R., Jin, Y., Echeverria, V., Milesi, M., Fan, J., Zhao, L., Alfredo, R., Li, X., & Gašević, D. (2025). The effects of generative AI agents and scaffolding on enhancing students' comprehension of visual learning analytics. *Computers and Education*, 234, 105322.
- Yan, X., Yang, X., Jin, N., Chen, Y., & Li, J. (2025). A general AI agent framework for smart buildings based on large language models and ReAct strategy. *Smart Construction*, 2(1), 0004.
- Yang, E.-W., Waldrup, B., & Velazquez-Villarreal, E. (2025). AI-HOPE-TGFbeta: A Conversational AI Agent for Integrative Clinical and Genomic Analysis of TGF- β Pathway Alterations in Colorectal Cancer to Advance Precision Medicine. *AI*, 6(7), 137.
- Yang, E.-W., Waldrup, B., & Velazquez-Villarreal, E. (2025). Conversational AI agent for precision oncology: AI-HOPE-WNT integrates clinical and genomic data to investigate WNT pathway dysregulation in colorectal cancer. *Frontiers in Artificial Intelligence*, 8, 1624797.
- Yang, W., Li, H., & Lee, J. C.-K. (2025). Tailoring AI agents for early learning: The Creative Project Approach. *Computers and Education: Artificial Intelligence*, 9, 100473.
- Yoon, S., & Hwang, J. (2025). AI agent-based indoor environmental informatics: Concept, methodology, and case study. *Building and Environment*, 277, 112879.
- Yoon, S., Song, J., & Li, J. (2025). Ontology-enabled AI agent-driven intelligent digital twins for building operations and maintenance. *Journal of Building Engineering*, 108, 112802.

- Yu, J., Zhao, J., Miranda-Moreno, L., & Korp, M. (2025). Modular AI agents for transportation surveys and interviews: Advancing engagement, transparency, and cost efficiency. *Communications in Transportation Research*, 5, 100172.
- Zhang, J., Yang, Z., Zhang, M., Chen, H., Zhao, L., & Liu, C. (2025). INF-SLiM: Large-Scale Implicit Neural Fields for Semantic LiDAR Mapping of Embodied AI Agents. *Journal of Field Robotics*. Advance online publication.
- Zhang, S., Qian, Y., Yao, Z., Ni, Z., & Zhang, Y. (2025). From approach to avoidance: How AI agent cognitive and affective empathy elicits the uncanny valley effect. *Telematics and Informatics*, 101, 102313.
- Zhao, B., & Cao, S. (2025). Exploring Knowledge Mining Using Large Language Model Combined with Knowledge Base and AI Agent. *Documentation, Information and Knowledge*, 42(4), 88–101.
- Zhao, Y., Qian, W., Chen, Y., Wu, D., Luo, Y., Gao, C., Wu, K., & Liu, Z. (2025). Effect of an AI agent trained on a large language model (LLM) as an intervention for depression and anxiety symptoms in young adults: A 28-day randomized controlled trial. *Applied Psychology: Health and Well-Being*, 17(5), e70067.



© 2025 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).