

Evolution of Artificial Intelligence Models and Agents: A Survey and Future Research Directions

B. Sai Sathvik*, M. Mani Muneesh, Dr. P. Sumithabhashini

Holy Mary Institue of Technology and Science, Hyderabad, Telanagana, India.

ABSTRACT

Artificial Intelligence (AI) is all about building the systems that can observe the world carefully, think about it, learn through experience and complete the task accurately. In the last decade, progress in data, computing power and neural network architectures has produced powerful AI models, especially deep learning systems, transformers and large language models (LLMs). Building on these models, a new class of systems called AI agents can pursue goals autonomously by planning, using tools, remembering context and acting on their environment. This paper reviews the foundations of AI, the major types of AI models, the architecture and types of AI agents, their applications across sectors, and the challenges of safety, bias, privacy and governance that accompany them.

Keywords: Artificial Intelligence, Machine learning, Deep learning, Neural Networks, Transformers, Large Language Models, AI Agents, Multi-agent Systems

INTRODUCTION

Artificial Intelligence is the branch of computer science whose central idea is to make machines behave intelligently. In standard textbooks the AI is referred as the study of intelligent agents: systems that observe their environment and take actions that increase their chances of achieving goals. Its central topics include reasoning, knowledge representation, planning, learning, communication, perception and the ability to move and manipulate objects.

The amount of data generated by people and machines now far exceeds what humans can absorb and analyse. AI helps convert this data into decisions and giving better suggestions to problem solving. Three layers are useful for understanding modern AI: AI methods (the techniques), AI models (trained systems produced by those techniques) and AI agents (systems that use models to act autonomously toward goals).

2. Artificial Intelligence Methods

Machine Learning

Machine learning (ML) is an important core topic in artificial intelligence that lets machines learn from data rather than being explicitly programmed. Its common types are supervised learning (learning a

mapping from labelled inputs to outputs), unsupervised learning (finding structure in unlabelled data) and reinforcement learning (learning actions that maximize cumulative reward through trial and error). Self-supervised learning, where a model creates its own training signal from raw data, underlies most modern language models.

2.1 Deep Learning and Neural Networks

Deep learning is a part of Machine learning. The neural networks take the biological word “neurons” from human brain as inspiration that are organized in layers. Training adjusts the weights between neurons, commonly with back-propagation, so that the network produces the desired output. Deep learning uses many layers to learn hierarchical features and has driven advances in vision, speech and language.

2.2 Natural Language Processing (NLP)

NLP can be defined as the area of AI that deals with the human language. Applications of NLP are machine translation, voice assistants, IVR systems in call centres, grammar checking, search and summarization.

2.3 Computer Vision

Computer vision allows computer to understand from the images and videos. It is used in medical image

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

analysis, face detection which is being used for biometric analysis in few institutions and signature identification, quality inspection and autonomous driving.

2.4 Knowledge-Based Systems and Robotics

A knowledge-based system (KBS) separates a knowledge base (rules, frames or cases from human experts) from an inference engine that reaches conclusions. Robotics and automation combine sensing, planning and control so that machines can carry out repetitive or hazardous physical tasks.

3.AI Models

An Artificial Intelligence (AI) model is a mathematical function that learns patterns from data to perform tasks such as prediction, classification, and content generation. The AI model development lifecycle

consists of several stages, including data collection and preprocessing, model selection, training, validation, testing, deployment, and monitoring. Initially, data is collected, cleaned, and transformed to improve its quality. An appropriate algorithm is then selected and trained to learn patterns by optimizing its parameters. Validation and testing are performed using separate datasets to evaluate model performance through metrics such as accuracy, precision, recall, and F1-score. After achieving satisfactory results, the model is deployed in a real-world environment. Continuous monitoring and periodic retraining help maintain its accuracy, reliability, and adaptability to changing data. Thus, a systematic AI development lifecycle ensures efficient, robust, and reliable model performance. Following table describes the various models, key idea in it and applications

Table 1: Performance Comparison of Different Models

Model type	Key idea	Typical use
Classical ML (regression, decision trees, SVM, k-means)	Statistical learning on structured data	Credit scoring, forecasting, clustering
Convolutional Neural Networks (CNN)	Learn spatial features with filters	Image classification, medical imaging
Recurrent Networks (RNN, LSTM)	Process sequences step by step	Speech, time series
Transformers	Self-attention over whole sequences	Language, vision, code, multimodal tasks
Large Language Models (LLMs)	Transformers trained on very large text corpora	Chat, writing, coding, reasoning, tool use
Generative models (GANs, diffusion)	Learn a data distribution to create new samples	Image, audio and video generation
Reinforcement learning models	Learn policies from reward	Games, robotics, control

3.1 Foundation Models and LLMs

Foundation models are large AI models that are trained on very big and large data sets. The transformer architecture, introduced in the year 2017, replaced recurrence with self-attention and made it practical to train very large models. Foundation models are pretrained on broad data and then adapted to many tasks through fine-tuning, prompting, retrieval augmentation or alignment with human feedback. Modern models are increasingly multimodal, handling text, images, audio and code together.

3.2 Evaluating Models

Models are assessed with metrics such as accuracy, precision, recall and F1 for classification, and with benchmarks and human evaluation for generative

systems. Important qualities besides accuracy include robustness, fairness, explainability, latency and cost.

4.AI Agents

An AI agent is a system that can work towards a goal by observing the information provided, making decision and taking actions. Whereas a model simply maps an input to an output, an agent runs in a loop: observe, reason, act, observe the result and repeat until the goal is met.

Agent loop: Goal → Perceive (inputs, tools, memory) → Reason & Plan → Act (call tool / API / write code) → Observe result → Update memory → repeat or finish.

4.1 Components of an Agent

- **Model (the “brain”)**: it is an LLM that interprets goals and decides next steps.
- **Perception**: text, images, sensor data or system state supplied to the agent.
- **Memory**: short-term context and long-term stores (databases, vector search) that keep information across steps.
- **Planning and reasoning**: breaking a goal into sub-tasks and revising the plan when results change.
- **Tools** : includes web search, code execution, databases, APIs, software interfaces or physical robots.
- **Guardrails and oversight**: permissions, validation and human approval for sensitive actions.

4.2 Types of Agents

Classical AI describes five kinds of agents: simple reflex agents (act on current input by rules), model-based agents (keep an internal state of the world), goal-based agents (choose actions that reach a goal), utility-based agents (choose actions that maximize expected benefit) and learning agents (improve from experience). Modern LLM-based agents combine several of these features, using reasoning-and-acting patterns in which the model alternates between thinking and calling tools.

4.3 Multi-Agent Systems

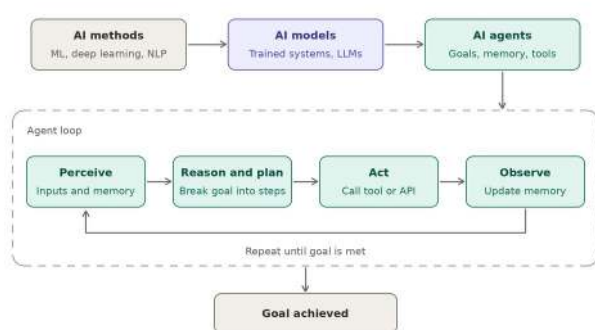


Figure 1: Multi Agent System

The Multi Agent systems contains multiple AI agents can cooperate, with roles such as planner, researcher, coder and reviewer, or compete, as in game-theoretic settings. Multi-agent designs can split complex work and cross-check results, but they add cost, coordination difficulty and new failure modes.

5.Applications of AI

1. **Healthcare**: AI in healthcare supports medical research and documentation, patient monitoring, medical diagnosis.
1. **Finance**: it helps in fraud detection in banks and organizations.
2. **Education**: many schools, institutes use digital classrooms, virtual classes.
3. **Transport and Automotive**: AI is used in traffic management, deciding suitable and short routes to reach destination. It is also used in autonomous self-driven cars.
4. **Agriculture**: in this sector it can be used for crop monitoring, yield production, crop selection.
5. **Data Security**: threat detection, malware and antivirus analysis, anomaly detection.
6. **E-commerce and Social Media**: recommendations, search, content moderation and trend analysis.
7. **Robotics and Manufacturing**: intelligent automation, predictive maintenance and quality control.
8. **Software Engineering**: AI tools can help us to generate code easily without any hardwork.
9. **Customer Service and Business Operations**: virtual assistants and agents that handle queries, schedule work and process documents.
10. **Science and Astronomy**: analysing large datasets, simulation and hypothesis generation.
11. **Entertainment and Music**: AI is used in games, generating images, helps in preparing music playlist according to the taste of the listener.

6. Challenges and Risks

- **Bias and fairness**: models can reproduce biases present in training data.
- **Hallucination and reliability**: generative models may produce fluent but incorrect output; errors can compound over many agent steps.
- **Privacy and data protection**: AI depends on large volumes of data, which raises consent and security issues.
- **Security**: prompt injection, data poisoning, model theft and misuse of agent permissions.
- **Transparency**: deep models are hard to explain, which matters in medicine, law and finance.

- **Employment and society:** automation changes jobs and skills.
- **Energy and cost:** training and running large models consumes substantial computing resources.
- **Accountability:** deciding who is responsible when an autonomous agent causes harm.

7.Future of AI

AI is expected to become an important part of many software systems. Likely directions are more capable multimodal models, smaller and more efficient models that run on personal devices, agents that carry out long multi-step tasks with human supervision, and stronger methods for interpretability, evaluation and alignment. AI will be used to strengthen both defensive and offensive cyber operations, and the importance of data will grow with its appetite for training data. Prudent governance, standards for testing and human oversight will be essential so that the technology brings broadly shared safety and prosperity.

CONCLUSION

This paper has discussed the principles of Artificial Intelligence, the main kinds of AI models and the emerging role of AI agents. Models supply the ability to observe, predict and generate; agents add goals, memory, planning and tools, turning that ability into action. Together they can solve problems that humans cannot easily handle alone, but their benefits depend

on careful design, testing, security and responsible governance. So it is very much Important to handle it carefully and need constant monitoring. The development of this field will change many parts of society, and engineers, researchers and policymakers share the responsibility of guiding it.

REFERENCES

1. S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed., Pearson, 2020.
2. Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
3. A. Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems (NeurIPS), 2017.
4. T. Brown et al., "Language models are few-shot learners," NeurIPS, 2020.
5. R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed., MIT Press, 2018.
6. S. Yao et al., "ReAct: Synergizing reasoning and acting in language models," ICLR, 2023.
7. I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT Press, 2016.

HOW TO CITE: B. Sai Sathvik, M. Mani Muneesh, Dr. P. Sumithabhashini, Evolution of Artificial Intelligence Models and Agents: A Survey and Future Research Directions, Int. J. Sci. R. Tech., 2026, 3 (10), 1524-1527. <https://doi.org/10.5281/zenodo.23835068>