Proceedings of the 3rd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/145/2025.21923

The Prospect of Embodied Intelligence and the Opportunities Brought by Large Models

Jialan Wu

The Chinese University of Hongkong (Shenzhen), Shenzhen, China
124090680@link.cuhk.edu.cn

Abstract: There is an emerging trend transitioning from conventional intelligent artificial systems to embodied intelligence, which represents a systematic function integrated with a physical carrier. Unlike the previously abstract AI, embodied intelligence leverages theories and technologies from artificial intelligence, robotics, mechanical manufacturing, and design. It accomplishes specific tasks through interactions with the real world, thereby exerting a direct or indirect influence on the physical environment. New technology like Open AI Sora has marked a closer step for AGI (Artificial General Intelligence). With the continuous improvement of large model technology and deep learning, embodied intelligence has unprecedented opportunities to develop. Therefore, this paper will focus on the prospect of embodied intelligence and the chances brought by large models. This paper is written based on summarizing and discussing many previous papers, using the method of literature review and research. By taking advantage of large model, embodied AI has the prospect to make progress in interactions with humans, more accuracy in execution, and more fields.

Keywords: embodied AI, large model, robotics, robot learning, humanoid robot

1. Introduction

Currently, the development of AI has progressed through two fundamental phases: computational intelligence and perceptual intelligence. To date, most scholars have focused their research on these two domains. Specifically, large language models such as ChatGPT and DeepSeek utilize both computational intelligence (CI) and perceptual intelligence (PI) to address natural language processing challenges. Technologies like computer vision, environment sensing, and speech recognition and processing are grounded in PI. Meanwhile, advancements in computer vision, embodied action optimization, embodied cognition, and embodied perception have established a foundation for the application of AI in broader fields. By deploying agents in real-world environments to interact with the physical world, embodied agents can execute tasks such as grasping, manipulating, and moving, thereby achieving perception and recognition. This is referred to as 'Higher Level Intelligence' [1] in the book 'Embodied Artificial Intelligence'. The High Level Intelligence we aspire to today encompasses not only perception but also overall development. Consequently, constructing a system capable of a wide array of imitation behaviors is both significant and challenging. It is therefore important to categorize the various fields in which embodied agents will be applied. This paper argues that these fields include perception, motion control, cognitive modelling, learning and adaptablity, which will be discussed in detail in following article. Embodied AI has a diverse range of application fields, one of which is humanoid robotics. Humanoid robots can be

DOI: 10.54254/2755-2721/145/2025.21923

deployed in domestic settings, military fields, and industry, and their development is based on cross-disciplinary research. In the realm of embodied AI, cross-disciplinary collaboration is commonplace, with fields such as electromechanics, deep learning, computer vision, sensing, and cognition all playing crucial roles. Large models, as a foundational component of these fields, can significantly impact embodied intelligence. This paper will focus on these aspects to enhance the understanding of the research content and prospects of embodied AI and to identify potential opportunities for its advancement.

2. Research of Embodied AI

The research field of embodied AI can be quite broad. However, to be simple, the most fundamental parts are perception and motion control, cognitive modelling, and learning and adaptability.

2.1. Perception and Motion

2.1.1. Perception

The primary objective of embodied perception is to empower machines to proactively comprehend their surrounding environment. This perception can be categorized into two distinct components: perception pertaining to humans and perception pertaining to objects [2].

2.1.1.1. Perception for human

Robot perception of people involves the use of sensors and algorithms by robots to recognize, understand, and respond to human behavior, states, and intentions. This perceptual ability forms the foundation for human-robot interaction, which encompasses visual perception, auditory perception, tactile perception, and intention perception. Among these, intention perception is particularly challenging, as it cannot be addressed solely through CI and PI. For instance, consider a rescue robot stationed by a swimming pool, tasked with saving individuals who are drowning. Before taking action, the robot must ascertain whether a person is genuinely drowning or merely engaging in a playful prank with friends. This necessitates the consideration of numerous factors, and robots are expected to emulate the decision-making processes of a real rescuer, possessing skills in rescue techniques, adaptability, and precision. Human perception is a crucial component due to the growing demand for seamless interaction between humans and robots.

2.1.1.2. Perception for object

Perception of objects refers to the robot using sensors and algorithms to identify, locate, understand, and operate objects. Perception can be divided into many parts, which are visual perception, tactile perception, and motion perception.

(1) Visual Perception

Visual perception includes object recognition, object detection, object segmentation, pose estimation, and texture and color recognition [3]. Machine learning (ML) involves training algorithms to learn patterns from data. In visual perception, ML techniques are used for tasks like object recognition, classification, and detection. For example, classify an image as containing a "cat" or "dog" or algorithms like SIFT (Scale-Invariant Feature Transform) and HOG (Histogram of Oriented Gradients) are used to identify key features in images.

(2) Tactile Perception

There are some basic properties of objects like mass, volume, inertia, coefficient of friction, degree of hardness and the shape of objects. These are essential factors needing consideration in the real world. By knowing these figures, embodied perception can perceive more abundant information such

DOI: 10.54254/2755-2721/145/2025.21923

as physical properties and geometric structures, interact with the environment, and obtain more information in the process of interacting with the environment. For example, from an operational point of view, embodied perception can autonomously interact with objects with high degrees of freedom that have never been seen before.

(3) Motion Perception

Motion perception is composed of motion detection, motion tracking, motion prediction, and motion analysis. There are lots of techniques used in motion perception. For instance, techniques such as frame differencing, optical flow, and background subtraction are commonly employed for motion detection, while Kalman Filters, Particle Filters, and Multi-Object Tracking (MOT) are typically utilized for motion tracking. Deep neural networks (DNNs), a component of deep learning, are particularly effective in motion prediction. In summary, motion perception constitutes a fundamental capability for robots and computer vision systems, facilitating their ability to comprehend and respond to dynamic environments.

2.1.2. Motion

Compared to the perception level of embodied intelligence, motion is more in the field of mechanism and real-world physics. In classification, motion has locomotion, manipulation, articulated motion, whole-body motion.

Robot motion depends on several key technologies to achieve precise and efficient movement. These include motion planning, control, perception, and the study of kinematics and dynamics[4].

2.1.2.1. Motion Planning

Motion planning involves creating paths for robots to follow, often using algorithms like A* or RRT to avoid obstacles. In dynamic environments, real-time adjustments are necessary to handle unexpected changes.

2.1.2.2. Motion Control

Motion control ensures robots follow these paths accurately. Simple tasks might use PID control, while more complex tasks, like assembly, require advanced methods such as Model Predictive Control (MPC) or force control.

2.1.2.3. Motion Perception

Motion perception enables robots to comprehend their surroundings. Techniques such as Visual Odometry and SLAM (Simultaneous Localization and Mapping) facilitate the navigation and mapping of the environment by robots, while simultaneously allowing them to maintain awareness of their position.

2.1.2.4. Kinematics and Dynamics

Kinematics and dynamics provide the theoretical basis for understanding robot movement. Kinematics focuses on position and velocity, while dynamics deals with the forces and torques involved. Together, these technologies enable robots to perform tasks in industries like manufacturing, healthcare, and autonomous driving.

DOI: 10.54254/2755-2721/145/2025.21923

2.2. Cognitive Modelling

Embodied intelligence emphasizes that the agent realizes intelligent behavior through the interaction of its body with the environment. Cognitive modeling refers to the simulation of human or animal cognitive processes through computational models, including perception, learning, memory, decision making, etc. The combination of embodied intelligence and cognitive modeling provides a new perspective for the design and research of agents.

The goal of cognitive modeling is to understand the nature of intelligence by using computational models to simulate cognitive processes in humans or animals. Though intelligence is a broad idea, this paper attempts to narrow it to intelligence for embodied cognition. That is “Intelligence is the result of a combination of body, environment and brain”.

The main method used in reinforcement learning includes symbolic models, connectionist models, behaviorist models, dynamic systems models [5]. However, the true challenge of cognitive modelling is no lack of models but is lack of enough statistic to test it.

Cognitive modeling will play a greater role in understanding human intelligence and developing smarter systems. By embedding cognitive models into embodied agents, more natural and efficient intelligent behavior can be achieved.

2.3. Learning and Adaptability

Learning and adaptability are the main abilities in robotics, which is used to enable robots to constantly learn and adapt to complex, dynamic task scenarios through interaction with the environment. Learning refers specifically to robot learning, which is different from DL or ML. Robot learning is the process by which a robot learns through interaction with the environment or from data to improve its behavior, decision-making, or control strategies while machine learning is a subfield of computer science and deep learning is a subfield of machine learning. Robot learning includes Reinforcement Learning (RL), Imitation Learning, Online Learning, and Transfer Learning. The robot learns to complete a specific task by interacting with the environment through trial and error, obtaining reward signals and optimizing its strategy. Environmental adaptability requires robots to be able to sense changes in the environment. This adaptability enables robots to apply new knowledge within novel environments.

3. Application of Embodied AI

The most classical application for embodied AI is humanoid robot [6]. Humanoid robots, designed to resemble and mimic human form and behavior, represent one of the most challenging areas of robotics. These robots are built with a body like humans, including a head, torso, arms, and legs, enabling them to perform tasks in environments designed for people. From assisting in households to working in disaster zones, humanoid robots have the potential to change how humans interact with machines.

In 1972, the Kato Laboratory of Waseda University in Japan developed the world's first full-size humanoid robot WABOT-1[37], but there was a large room for improvement in its movement ability. In 2008, NAO, a small humanoid robot developed by Aldebaran, a French company, realized commercial landing in the educational scene and promoted the industrialization of humanoid robots.[2]

Nowadays, the Optimus humanoid robot developed by Tesla introduces the visual processing system of intelligent car driving into the humanoid robot to improve the perception and operation ability of the humanoid robot. Several humanoid robots, robot dogs, and other products released by Yushu Technology have been sold, playing a significant role in advancing the research and application of embodied intelligence and humanoid robotics.

DOI: 10.54254/2755-2721/145/2025.21923

The future of humanoid robots is both exciting and uncertain. Companies like Boston Dynamics, Tesla, and Softbank are already developing humanoid robots for commercial use. However, widespread adoption will depend on overcoming technical challenges, reducing costs, and addressing societal concerns. As humanoid robots become more integrated into daily life, they have the potential to transform industries, improve quality of life, and redefine the relationship between humans and machines.

4. Opportunities and Challenges Brought by Large Models

4.1. Opportunity

4.1.1. Natural language interaction

Natural language interaction facilitates more intuitive communication between robots and humans. Large language models, such as ChatGPT, significantly enhance the efficiency of this communication.

4.1.2. Smarter decision making

Robots with advanced decision-making capabilities can analyze complex situations and choose the best course of action based on available data.

4.1.3. More accurate execution

Precision and accuracy are critical for robots performing tasks that require fine motor skills or exact movements. With improved control algorithms and high-resolution sensors, robots can execute tasks with greater precision.

4.2. Challenge

At present, embodied intelligence has shown its application in many scenarios. However, the application of embodied intelligence also faces problems.

4.2.1. Real-Time Interaction

Embodied intelligence requires robots to process sensory data and make decisions in real-time. This is challenging because physical environments are dynamic and unpredictable. For example, a robot navigating a crowded space must constantly adjust its path to avoid obstacles while maintaining smooth movement.

4.2.2. Lack of large-scale real data

The lack of data accumulation in various scenarios and domains prevents data-driven learning methods from being directly transferable to the study of embodied intelligence.

4.2.3. Energy Efficiency and Hardware Limitations

Embodied systems rely on physical hardware, which has limitations in power, durability, and computational capacity. For instance, a small robot might not have enough battery life to complete a long task, or its sensors might fail under extreme conditions.

DOI: 10.54254/2755-2721/145/2025.21923

4.2.4. Safety and Ethical Concerns

As robots interact more closely with humans, ensuring their safety and ethical behavior becomes crucial. A robot working in a home or hospital must avoid harming people, even if it means sacrificing efficiency. Additionally, questions about privacy, data security, and the ethical use of autonomous systems add layers of complexity to the development of embodied intelligence.

5. Conclusion

Embodied AI represents a significant leap in the evolution of artificial intelligence, moving beyond computational and perceptual intelligence to integrate physical interaction with the environment. By embedding agents into real-world scenarios, embodied intelligence enables machines to perform complex tasks such as grasping, manipulating, and navigating, while also achieving advanced perception and cognition. This shift toward "higher-level intelligence" emphasizes the importance of systems capable of imitation, adaptability, and interaction, which are crucial for applications ranging from humanoid robots to industrial automation.

The research of embodied AI is built on several foundational pillars: perception and motion control, cognitive modeling, and learning and adaptability. Perception allows robots to understand their environment and interact with humans and objects, while motion control ensures precise and efficient physical actions. Cognitive modeling bridges the gap between physical interaction and intelligent decision-making, enabling robots to simulate human-like reasoning. Learning and adaptability empower robots to improve their performance over time, adapting to new tasks and environments through techniques like reinforcement learning and transfer learning.

Despite its potential, embodied AI faces significant challenges, including real-time interaction, energy efficiency, and ethical concerns. The integration of large models, such as ChatGPT and other sophisticated AI systems, presents promising avenues for enhancing natural language interaction, decision-making precision, and execution accuracy. Nevertheless, the scarcity of large-scale real-world data and the inherent limitations of current hardware pose significant obstacles to their widespread adoption.

As embodied AI continues to evolve, it holds the potential to revolutionize industries, improve quality of life, and redefine human-robot collaboration. By addressing current challenges and leveraging advancements in AI, robotics, and cognitive science, embodied intelligence can pave the way for a future where machines seamlessly integrate into our daily lives, performing tasks with human-like precision and adaptability.

References

[1] *Pfeifer, R., & Iida, F. (2004). Embodied artificial intelligence: Trends and challenges. Lecture notes in computer science, 1-26.*
[2] *Zhang Weinan & Liu Ting. Research and application of embodied intelligence. Journal of Intelligent Systems, 1-9.*
[3] *Wade, N., & Swanston, M. (2013). Visual perception: An introduction. Psychology Press.*
[4] *Brady, M. (Ed.). (1982). Robot motion: Planning and control. MIT press.*
[5] *Busemeyer, J. R., & Diederich, A. (2010). Cognitive modeling. Sage.*
[6] *Kuniyoshi, Y., Yorozu, Y., Ohmura, Y., Terada, K., Otani, T., Nagakubo, A., & Yamamoto, T. (2004, July). From humanoid embodiment to theory of mind. In Embodied Artificial Intelligence: International Seminar, Dagstuhl Castle, Germany, July 7-11, 2003. Revised Papers (pp. 202-218). Berlin, Heidelberg: Springer Berlin Heidelberg.*