

---

# Self-Identification in AI: ChatGPT's Current Capability for Mirror Image Recognition

---

[Berend Watchus](#) \*

Posted Date: 15 November 2024

doi: 10.20944/preprints202411.1112.v1

Keywords: Self-Recognition; Artificial Intelligence (AI); ChatGPT; Mirror Test; Mirror Image Recognition; TurtleBot 3; Image Processing; Computer Vision; Convolutional Neural Networks (CNNs); YOLO (You Only Look Once); Reflections in Art; Symbolic Interpretation; AI and Self-Awareness; Emergent Self-Awareness; Feedback Loops; Psychological Self-Perception; Robot Behavior; Symbolism in Mirrors; Human-Robot Interaction; Real-Time Self-Recognition; Identity and Reflection; Visual Data Analysis; Artistic Perspective in AI; Artificial Consciousness; AI-Driven Contextual Analysis; Image Distortion (Diagonal Reflection); Human-Like Identity Perception; AI and Emotion; Feedback Confirmation Mechanisms; Pre-programmed Robot Models



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Concept Paper

# Self-Identification in AI: ChatGPT's Current Capability for Mirror Image Recognition

Berend F. Watchus

Affiliation: mailonlinebw@protonmail.com)

**Abstract:** This paper explores a novel approach to self-recognition in artificial intelligence (AI) systems by integrating ChatGPT with a mirror image recognition process. Unlike previous models that emphasize emergent self-awareness through complex feedback loops and emotional interfaces, this research presents a more direct method of self-identification, where ChatGPT plays a key role in interpreting visual data from the robot's reflection. We detail the steps involved in equipping the TurtleBot 3 with the ability to recognize itself via image processing and how ChatGPT aids in confirming self-awareness. This research bridges current AI capabilities and the long-standing mirror test (Gallup, 1970) in an accessible, technical manner.

**Keywords:** self-recognition; artificial intelligence (AI); ChatGPT; mirror test; mirror image recognition; TurtleBot 3; image processing; computer vision; convolutional neural networks (CNNs); YOLO (You Only Look Once); reflections in art; symbolic interpretation; AI and self-awareness; emergent self-awareness; feedback loops; psychological self-perception; robot behavior; symbolism in mirrors; human-robot interaction; real-time self-recognition; Identity and reflection; visual data analysis; artistic perspective in AI; artificial consciousness; AI-driven contextual analysis; image distortion (diagonal reflection); human-like identity Perception; AI and emotion; feedback confirmation mechanisms; pre-programmed robot models

---

## Introduction

The mirror test has long been used as a measure of self-awareness in animals (Gallup, 1970). This test, while historically a measure of physical self-recognition, also has implications for understanding sentience in AI. Recent advancements in artificial intelligence have made it possible to explore this concept in the context of AI systems, specifically robots. The question of whether a machine can recognize its own reflection challenges foundational ideas in AI, consciousness, and the self-recognition process.

Previous work by the author (Watchus, 2024) focused on emergent self-awareness through complex feedback loops and interactions in AI systems, exploring the role of the insula and affective interfaces. These earlier studies investigated how AI could develop a form of self-awareness through emotional feedback loops, where feedback from the robot's environment could lead to emergent patterns of self-reflection. This paper presents a simpler, more direct approach to self-recognition in AI by incorporating ChatGPT into a robot's ability to recognize its own reflection in a mirror. This approach avoids the emergent complexity of earlier models and instead leverages current image recognition technologies and AI-driven contextual analysis.

## Related Work

Self-awareness in artificial intelligence is a deeply explored topic, with various models examining feedback loops, emotional interfaces, and more. The mirror test has been pivotal in measuring self-recognition in animals (Gallup, 1970), providing a foundation for research into AI self-recognition. However, traditional approaches to AI self-awareness often rely on feedback loops and emergent processes that require significant computational complexity (Miller, Suddendorf, & Povinelli, 2005).

In contrast, this paper introduces a method that combines modern image processing with ChatGPT to perform immediate self-recognition. By focusing on image features such as shape, size,

and contextual comparisons, this research integrates AI and computer vision to perform self-recognition in a manner that is more accessible and immediate than previous, more complex systems.

## Methods

**Integration of ChatGPT for Mirror Image Recognition:** The process of self-recognition using ChatGPT involves a structured pipeline, beginning with the TurtleBot 3's visual sensors and concluding with a feedback loop confirming the robot's awareness of its own reflection.

**Visual Input Capture and Processing:** The TurtleBot 3 is equipped with a camera and visual sensors that capture an image of the robot's reflection in real-time. This raw image data is passed to an image-processing module equipped with advanced computer vision algorithms. Specifically, Convolutional Neural Networks (CNNs) and YOLO (You Only Look Once) are employed for image analysis. These algorithms are particularly adept at object detection and image segmentation, enabling the system to identify and locate the robot's reflection in the mirror.

**ChatGPT's Art and Reflection Analysis:** Beyond the robot's direct visual recognition of its reflection in a mirror, ChatGPT's understanding of reflections in artwork provides a nuanced approach to mirror image recognition. Specifically, ChatGPT interprets not only basic reflections but also people depicted in paintings and pictures standing in front of or diagonally reflected in mirrors. This goes beyond simple object recognition to involve a layered understanding of artistic perspective, symbolism, and psychological themes.

In works where subjects are shown in front of or at an angle to a mirror, ChatGPT recognizes the angular distortion caused by the diagonal reflection and adjusts its understanding of the relationship between the real-time subject and their mirrored image. This understanding is based on both visual cues (e.g., shape, orientation, symmetry) and the symbolic context often employed in art. For example, ChatGPT interprets the reflection in a mirror as not only a reversal of left and right but also as an expression of identity, psychological conflict, or narrative structure.

ChatGPT's role in recognizing these complex mirror relationships enhances the accuracy of the self-recognition process by ensuring that the robot correctly identifies distorted reflections, such as those viewed from a diagonal angle, and understands the deeper meaning of the reflection in its environment.

**Role of ChatGPT:** Once the image processing module has analyzed the reflection, the resulting features (such as key points, shape, and size) are passed to ChatGPT. While ChatGPT itself does not process visual data, it plays a crucial role in interpreting the results of the image recognition process. ChatGPT's task is twofold:

**Processing Context:** ChatGPT receives key image features, such as the robot's shape and size, and compares them against pre-programmed models of the robot's physical structure. This step allows ChatGPT to verify if the robot is indeed viewing its own reflection.

**Self-Recognition Confirmation:** If the visual input matches the robot's known parameters, ChatGPT confirms self-recognition. This confirmation triggers a feedback loop, where the robot's action modules respond to the recognition, such as avoiding the mirror or interacting with its reflection.

**Feedback Loop and Self-Recognition Behavior:** This process, though technically simpler than the emergent self-awareness models previously discussed (Watchus, 2024), offers a more direct approach. It highlights how ChatGPT can act as a conduit for self-recognition confirmation, without relying on complex emotional interfaces or adaptive feedback loops. Once the robot confirms it is looking at itself, its behavior can change accordingly, as seen in the mirror test (Gallup, 1970), though achieved through a more direct, technical method.

## Discussion

This research proposes a simple but effective method for self-recognition in robots by integrating ChatGPT with advanced image processing techniques. The approach contrasts with earlier models that explored emergent self-awareness through complex feedback systems (Watchus, 2024). In contrast, this research demonstrates that ChatGPT can be used as a direct interface for interpreting visual data and confirming self-recognition without the need for emotional feedback loops or

emergent behaviors. By directly comparing the features of the robot's reflection with pre-programmed models, the AI can determine whether it is viewing itself.

### **Expanded Role of ChatGPT in Understanding Reflections in Art:**

In addition to recognizing basic mirror images, ChatGPT's ability to understand more complex reflections—such as those of people depicted in front of or diagonally reflected in mirrors in artwork—represents a significant expansion in the capability of AI to interpret not just visual stimuli, but also the symbolic and psychological dimensions embedded in these images. This ability allows ChatGPT to engage with artistic representations that explore identity, self-awareness, and psychological conflict, much like how artists have used mirrors throughout history to challenge perceptions of the self.

In artworks where subjects are reflected diagonally in mirrors, ChatGPT understands the distortion caused by the angle and correctly identifies the relationship between the subject and their reflection, whether the subject is aligned directly in front of the mirror or standing at an oblique angle. This understanding is critical for applications where the robot must interpret distorted reflections accurately, and it also enriches the broader context of mirror testing in AI, as it extends beyond mere self-recognition to include an interpretation of reflective self-awareness and self-perception.

The integration of ChatGPT into such scenarios opens the door to more nuanced and dynamic self-recognition systems in robots, as the understanding of reflections in art and psychological self-awareness could inform future developments in robotic behavior, making it more sophisticated and adaptable to human-like experiences of identity.

### **Conclusions**

The integration of ChatGPT for mirror image recognition represents a significant step toward simplified self-awareness in AI systems. This direct method, in contrast to earlier complex feedback-based models, provides a more accessible and practical approach to self-recognition. By using advanced computer vision techniques and leveraging ChatGPT's processing capabilities, this research demonstrates that AI systems can begin to perform self-recognition tasks in real-time, which has broader implications for future AI development and interaction. The inclusion of artistic and psychological reflections further enhances the understanding of mirror recognition, positioning ChatGPT as a tool capable of interpreting both literal and symbolic mirror imagery, enriching the broader implications of self-awareness in AI.

### **References**

- Bohus, D., & Rudnicky, A. I. (2005). Dialog in a virtual reality-based multi-modal environment. Proceedings of the International Conference on Spoken Language Processing.
- Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam.
- Gallup, G. G. (1970). Chimpanzees: Self-recognition. *Science*, 167(3914), 86-87.
- Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication: Does the left hemisphere dominate the right? *Science*, 289(5472), 61-63.
- Miller, R., Suddendorf, T., & Povinelli, D. J. (2005). The mirror test and the evolution of self-recognition. *Animal Cognition*, 8(2), 177-187.
- Watchus, B. (2024). *The Unified Model of Consciousness: Interface and Feedback Loop as the Core of Sentience*. Preprints.org.
- Watchus, B. (2024). *Towards Self-Aware AI: Embodiment, Feedback Loops, and the Role of the Insula in Consciousness*. Preprints.org.
- Watchus, B. (2024). *Simulating Self-Awareness: Dual Embodiment, Mirror Testing, and Emotional Feedback in AI Research*. Preprints.org.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.