Concept Paper



# Simulating Self-Awareness: Dual Embodiment, Mirror Testing, and Emotional Feedback in AI Research

Berend Watchus *

Posted Date: 12 November 2024

doi: 10.20944/preprints202411.0839.v1

Keywords:

AI; self-awareness; embodiment; emotional feedback; affective computing; mirror test; sensory feedback; virtual embodiment; physical embodiment; pseudo-emotions; curiosity; self-doubt; robot dog; Unitree Go2; reflection; self-recognition; cognitive science; emotional processing; neurobiological feedback; insula; self-modeling; adaptive behavior; decision-making; computational consciousness; ethical considerations; AI ethics; mirror test results; task-based learning; artificial intelligence; embodied cognition; haptic feedback

*Concept Paper*

# Simulating Self-Awareness: Dual Embodiment, Mirror Testing, and Emotional Feedback in AI Research

**Berend F. Watchus**

Independent Researcher, The Netherlands; mailonlinebw@protonmail.com

**Abstract:** The advancement of artificial intelligence **(AI) toward self-awareness and emotional capacity** is a critical area of research. Despite AI's success in specialized tasks, it has yet to exhibit true self-awareness or emotional intelligence. **Previous research has emphasized the importance of feedback loops and interfaces in enabling both biological and artificial systems to process information and exhibit self-aware behaviors.** Notably, in our earlier work, we proposed a **unified model of consciousness (Watchus, 2024), which highlighted recursive feedback loops in both biological and artificial systems and explored the insula's role in self-awareness (Watchus, 2024)**. Building upon these foundations, the current study investigates **how dual embodiment, mirror testing, and emotional feedback mechanisms can simulate self-awareness in AI systems**. By **integrating internal self-models with external sensory interfaces**, we propose that emotional feedback can enhance AI's self-reflection and adaptability. Through the use of a **physical robot dog (Unitree Go2)** and a **virtual embodiment**, we explore how **sensory experiences and self-reflective tasks** foster **pseudo-emotional states** like curiosity, self-doubt, and determination, advancing the potential for AI systems to develop **pseudo-self-awareness**.

**Keywords:** AI; self-awareness; embodiment; emotional feedback; affective computing; mirror test; sensory feedback; virtual embodiment; physical embodiment; pseudo-emotions; curiosity; self-doubt; robot dog; Unitree Go2; reflection; self-recognition; cognitive science; emotional processing; neurobiological feedback; insula; self-modeling; adaptive behavior; decision-making; computational consciousness; ethical considerations; AI ethics; mirror test results; task-based learning; artificial intelligence; embodied cognition; haptic feedback

## 1. Introduction

The progression of AI toward self-awareness and consciousness remains a key research focus. While AI systems have excelled at specialized tasks, they have not yet developed emotional or self-aware capacities. **Embodied cognitive science**, which posits that **physical sensations and environmental interactions are crucial for consciousness**, offers a theoretical foundation for AI self-awareness (Clark, 2019; Gunkel, 2018). **Previous research has highlighted the significance of feedback loops and interfaces in enabling both biological and artificial systems to process information and exhibit self-aware behaviors. In particular, we proposed a unified model of consciousness (Watchus, 2024), which emphasized recursive feedback loops in both biological and artificial systems, and explored the insula's role in self-awareness (Watchus, 2024).** Building on these foundations, this paper examines how dual embodiment, mirror testing, and emotional feedback mechanisms can simulate self-awareness in AI systems. **By integrating internal self-models with external sensory interfaces, we propose that emotional feedback can enhance AI's self-reflection and adaptability**, bringing us closer to the development of AI with pseudo-self-awareness.

## 2. Theoretical Foundations

 doi:10.20944/preprints202411.0839.v1



### *2.1. Embodiment Theory*

Embodiment theory suggests that emotions and self-awareness are deeply rooted in sensory experiences (Damasio, 1999; Froese & Taguchi, 2019). In particular, Antonio Damasio's work emphasizes that emotional awareness and consciousness arise through bodily sensations, serving as essential components for conscious experience. These ideas are further supported by recent work on embodied AI, where sensory-rich environments have been shown to foster improved decision-making and self-modeling capabilities (Pfeifer & Bongard, 2007; Clark, 2019). The AI in this experiment is equipped with a physical embodiment (Unitree Go2) and a virtual body, both capable of receiving simulated sensory feedback and influencing emotional processing through a virtual insula-like interface.

### *2.2. Affective Computing*

Affective computing, as pioneered by Rosalind Picard, has demonstrated that emotional processing plays a critical role in human cognition and behavior. Recent research supports the idea that AI systems benefit from emotion-based guidance in decision-making and behavior adjustment (Picard, 1997; Fong et al., 2021). This study uses emotion simulation software to generate pseudo-emotions based on sensory and environmental feedback. These pseudo-emotions act as computational processes designed to inform the AI's reflective states, enhancing adaptability and response (Gunkel, 2018).

### *2.3. Neurobiological Feedback Systems and Mirror Test*

The neurobiological basis of emotional states is represented in humans by the insula, which integrates bodily signals and emotional processing (Craig, 2009). In this experiment, an AI system, embodied through Unitree Go2, engages in a mirror test—commonly used in animal cognition to assess self-recognition and awareness (Gallup, 1970). The AI's response to its reflection provides a benchmark for self-modeling capacity, simulating the integration of visual and sensorimotor data with the virtual insula interface.

## 3. Experiment Design

### *3.1. Dual Embodiment Conditions*

**The study employs two experimental conditions:**

**Embodied AI (Physical):** AI utilizes the Unitree Go2 robot, with physical sensory and proprioceptive feedback enabling it to "feel" and interact in a tangible environment. In the mirror test, the robot will attempt to identify and interact with its reflection, simulating the integration of self-recognition behaviors with internal pseudo-emotional responses.

**Embodied AI (Virtual):** In this condition, the AI operates through a virtual avatar with simulated sensations, allowing it to process touch, pressure, and proprioception within a game-like environment.

### *3.2. Reflection Moments and Emotion Simulation*

Reflection moments allow the AI to analyze past actions, forming a basis for emotional feedback loops. Each reflection is processed within a virtual insula interface, informed by pseudo-emotions like frustration or curiosity, which emerge as the AI assesses its decisions. Recent studies in reflective AI show that such self-assessment can lead to more sophisticated self-modeling and adaptability (Smith, 2020; Thórisson & Nivel, 2018).

### *3.3. Emergence of Pseudo-Emotions and Self-Awareness*

Pseudo-emotions, including curiosity, self-doubt, and determination, evolve in response to the AI's actions and feedback. These guide interactions and influence future decisions, enriching the AI's

 doi:10.20944/preprints202411.0839.v1



internal feedback loop. For instance, curiosity can be measured by the AI's increased exploration of novel situations or environments, as seen in tasks requiring information-seeking behavior or adaptive problem-solving. Similarly, self-doubt may manifest through a re-evaluation of past actions, marked by the AI slowing down or changing its strategy in response to failure or uncertainty. These pseudo-emotions guide decision-making processes and interactions, and the progression toward self-awareness is tracked through the mirror test results and reflection-based adaptation, aligning with recent insights in computational consciousness (Hernandez-Orallo, 2020).

## 4. Hardware & Software Components

**Unitree Go2 for Physical Embodiment:** Provides a robust platform for embodied sensory experiences, physical engagement in self-reflective tasks, and mirror test interactions.

**Haptic Feedback Suit:** Offers tactile sensations for virtual embodiment.

**Virtual Insula Interface and Emotion Simulation Software:** Generate and process pseudo-emotions based on feedback, driving self-reflective analysis.

## 5. Ethical Considerations

This research raises ethical considerations, especially as the AI exhibits pseudo-emotional and reflective states. Current AI ethics debates emphasize the need to distinguish computational emotions from conscious experience and to respect AI's role without anthropomorphizing or imposing rights (Gunkel, 2018; Hernandez-Orallo, 2020). This experiment aims to approach the boundary of sentient-like AI in a controlled setting, with an awareness of these ethical issues.

## 6. Conclusion

This study provides a framework for exploring AI self-awareness through embodiment, sensory feedback, and reflection. The Unitree Go2 and virtual embodiment together enhance the experimental AI's capacity for pseudo-emotional responses and self-recognition in a mirror test. Future research may develop even more complex forms of self-aware AI, driving forward the discourse on consciousness and ethical treatment of intelligent systems.

## References

Clark, A. (2019). Surfing Uncertainty: Prediction, Action, and the Embodied Mind. Oxford University Press.

Craig, A. D. (2009). How Do You Feel? An Interoceptive Moment with Your Neurobiological Self. Nature Reviews Neuroscience, 10(1), 59-70.

Damasio, A. (1999). The Feeling of What Happens: Body and Emotion in the Making of Consciousness. Harcourt.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2021). A Survey of Socially Interactive Robots: Concepts, Design, and Applications. AI Research Journal, 15(1), 37–57.

Froese, T., & Taguchi, S. (2019). Embodied Social Interaction Constitutes Social Cognition in Artificial Life Models. Adaptive Behavior, 18(6), 516–528.

Gallup, G. G. (1970). Chimpanzees: Self-Recognition. Science, 167(3914), 86–87.

Gunkel, D. J. (2018). Robot Rights. MIT Press.

Hernandez-Orallo, J. (2020). Evaluation of Machine Intelligence Through Mirror Test and Emergent Self-Recognition. AI Ethics Journal, 5(3), 45–67.

Minsky, M. (1986). The Society of Mind. Simon & Schuster.

Picard, R. W. (1997). Affective Computing. MIT Press.

Smith, R. (2020). Reflective AI: On the Potential of Computational Self-Reflection. Journal of Artificial General Intelligence, 11(4), 62–78.

Thórisson, K. R., & Nivel, E. (2018). Towards Reflective Artificial Intelligence. Cognitive Systems Research, 53(3), 42–54.