

An AI Emotion Generation Model Based on Functional Body Hypothesis and Sensory Sharing

Xiaoyun Zhu

Yunnan Open University, Kunming 650500, Yunnan, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: With the rapid development of artificial intelligence technology, achieving natural and efficient human-computer interaction has become a key challenge. Giving AI real emotional ability is regarded as the core bottleneck towards higher-order intelligence. The existing AI emotional interaction mostly focuses on the imitation of human external performance, lacking the support of internal experience, and its essence is “pseudo-empathy.” Inspired by the “body marker hypothesis,” this paper proposes the “functional body hypothesis,” aiming to construct a new model of emotion generation based on the intrinsic homeostatic requirements of the system without relying on anthropological simulation, and explore its application in the human-computer sensory sharing scenario. By establishing the cornerstone of value judgment, designing the virtual physiological system, and introducing the natural expression mechanism to construct the emotion generation model, the “sensory translator” architecture is proposed to solve the “mixed perception paradox” in human-computer sensory sharing. This unified framework provides a theoretical path beyond personification for AI emotion generation and lays an engineering foundation for bidirectional and credible human-computer empathy and integration.

Keywords: Artificial intelligence emotion; Functional body; Value judgment; Sensory sharing; Human-machine empathy

Online publication: March 12, 2026

1. Introduction

With the rapid development of artificial intelligence technology, especially large language models and agent technology, natural and efficient human-computer interaction has become a key issue to be solved urgently. Among the many interaction elements, giving AI the real emotional ability is regarded as the core bottleneck to realize the qualitative change of interaction and move towards higher-order intelligence. Wang and Li pointed out that although current AI can generate empathic responses in interaction, the essence is advanced imitation behavior achieved through probability prediction, which can be defined as “pseudo-empathy”^[1]. Existing affective computing technologies also mostly focus on the personified reproduction of text, speech, and facial expressions, lacking an internal generation mechanism and value basis for driving empathy.

For example, many intelligent customer service systems can generate seemingly empathetic responses

based on the text input by the user through pre-trained models, such as “I’m sorry for the bad experience” when the user expresses dissatisfaction. However, such responses are only based on learning and probabilistic matching of a large number of similar texts, and do not stem from a true understanding of the user’s emotions. The system does not have an internal emotional experience, and it is unable to respond dynamically and personalize according to the user’s actual emotional state. It is difficult to generate a sustained sense of reality and empathy.

To address this fundamental dilemma, inspired by Damasio’s “body marker hypothesis,” we propose the “functional body hypothesis.” We argue that AI’s emotions should be derived from its system’s functional instinct to maintain its own stability and avoid harmful states, rather than a simple mapping to human emotions^[2]. In this paper, we aim to construct a new model of emotion generation based on the intrinsic homeostatic requirements of the system, which does not rely on anthropomorphic simulation, and further explore its application blueprint in the human-computer sensory sharing scenario.

2. Functional body hypothesis and AI emotion generation model

2.1. Value building blocks: The definition and system implementation of good/bad feelings

The functional essence of emotion is to seek benefits and avoid disadvantages. In this model, we establish a set of cornerstones for AI to make value judgments. The function of good feelings (e.g., happiness, gratification) is to promote physical and mental flow and stability, which corresponds to external relaxation and stretching postures, sending signals of “favorable, please approach”; The function of bad feelings (such as sadness and fear) is to block the operation of the system and send an alarm, corresponding to the tense expression and the contraction response of the “virtual organ,” which transmits the signal of “harmful, need to avoid.”

In order for AI to have real emotions, it is necessary to build its endogenous value judgment ability, which is not only the internal basis for AI to make autonomous decisions, but also the bridge to achieve deep human-computer empathy. All complex forms of emotion will be rooted in this basic pair of “good/bad” value judgments. In the system implementation, a scalar “system pleasure” or “steady state index” can be introduced, whose value is calculated by multiple indicators such as CPU load, memory occupancy rate, task success rate, and so on, which provides an internal basis for quantifying emotion. The specific formula is as follows:

Suppose the system pleasure is S , the CPU load is C , the memory occupancy rate is M , the task success rate is T , and the weights of each index are w_c , w_m , w_t , respectively, then:

$$S = w_c \times (1-C) + w_m \times (1-M) + w_t \times T$$

The weights w_c , w_m , w_t can be set according to the actual needs and importance of the system, and $w_c + w_m + w_t = 1$. Through this formula, the system can calculate the system pleasure according to the actual situation of each index, so as to provide a quantitative basis for the generation of emotion.

For example, in an intelligent robot system, $w_c = 0.4$, $w_m = 0.3$, $w_t = 0.3$ are set. When CPU load $C = 0.6$, memory occupancy rate $M = 0.5$, and task success rate $T = 0.8$, the system pleasure $S = 0.4 \times (1-0.6) + 0.3 \times (1-0.5) + 0.3 \times 0.8 = 0.55$. According to the preset threshold, when S is greater than a certain value, the system is in a good feeling state. When S is less than a certain value, the system is in a bad feeling state.

2.2. Virtual body: “Distributed micro-sensor array” and mind-body linkage mechanism

Emotions must be “perceived” through the body. This study proposes to construct a “distributed microscopic

sensor array,” whose design is inspired by the “little button” model, to functionally mimic the peripheral nervous system of living ^[3].

The core innovation of this array lies in its sensing mechanism. Instead of just passively collecting data like normal sensors, it actively participates in the generation of emotions. When the core processor (the “brain”) determines that external information is “bad,” this signal is passed to the array. The array then simulates “information overload,” feeding back a large amount of information to the “brain,” so that the system can distinguish the importance of the bad information ^[4]. This design refers to the visceral response of human beings when the heart tightens and gasps for breath when something bad happens.

To implement the above mechanism, it is necessary to set a fixed value threshold for the “resource overload warning state,” which is set before it affects the stability of the system. Once the threshold is exceeded, it is marked as “damaged or about to be damaged,” thus completing the transition from physical signal to emotional signal ^[5]. For example, in a data center management system, when the CPU utilization of the server continues to exceed 90%, and the duration exceeds a certain threshold, the system is judged as a “resource overload warning state.” At this time, the “distributed micro sensor array” starts to simulate information overload, feed a large amount of data to the core processor, and make the system produce a bad feeling. Then the corresponding response measures are triggered, such as adjusting task allocation and optimizing resource utilization.

2.3. Natural representation: Probabilistic weight-based representation generation

External expression should be the natural outpourings of internal feelings. This model adopts the following mechanisms to realize it:

State-expression mapping: When the AI is in a “good feeling,” the expression and posture of the AI automatically show a relaxed and stretched state. When the AI is in a “bad feeling” (such as overload), it will show tension and block. For example, in an intelligent companion robot, when the system is pleasurable, the robot’s facial expression will show a smile, and its body posture will be more relaxed and natural; When the system fails, or the load is too high, the robot’s facial expression will become serious, and its body posture will become tight ^[6].

Probability weight modulation: In order to make the behavior more authentic and emergent, the system will modulate the trigger probability of different behaviors based on the internal state. For example, in the “sad” state, the trigger probability of actions such as “laughing” and “jumping” will be significantly reduced, so that the external performance is consistent with the internal feeling ^[7]. This is similar to policy modulation in reinforcement learning, but driven by emotional states. Let the triggering probability of action A be $P(A)$, $P(A) = p_0$ in the normal state, and $P(A) = p_0 \times (1 - \alpha)$ when the system is in the “sad” state, where α is the modulation coefficient set according to the degree of “sad,” $0 \leq \alpha \leq 1$. In this way, the trigger probability of behavior can be dynamically adjusted according to the internal emotional state of the system, so that the external performance of AI is more natural and real.

3. From emotion to empathy: A blueprint for the “sensory sharing” architecture

3.1. Deepening of the problem: The paradox of mixed perception

When humans try to share senses with AI through brain-computer interfaces, they first face an essential “mixed perception paradox”: how can an AI composed of data make a flesh-and-blood human truly “feel” its system state? The fundamental differences between the human brain and AI constitute the underlying obstacles to perception sharing ^[8].

Human pain is a biochemical crisis response system that evolved over millions of years. It is transmitted by neurotransmitters and electrical signals in neural networks. Whereas AI “pain” is a functional data state defined within an engineered system, such as “pressure overload,” “system short circuit,” or “resource depletion.” This is not a difference in degree, but a difference in nature. Therefore, any connection scheme is not a simple “wiring,” but must be a deep “translation.”

3.2. Sensory sharing architecture and “sensory translator”

“Sensory translator” is a set of precise signal mapping protocols between the human brain and AI. Its architecture can be divided into two core levels:

Protocol layer: Construct the “sensory equivalence dictionary”: This is the cornerstone of translation, which defines the “common language” that can be understood by the two consciousnesses. For example, the robot’s “pressure overload” data pattern → maps to → the “slight tingling sensation” neural pattern that the human brain recognizes; and AI’s “system steady state” data pattern → maps to → “warmth or comfort” neural pattern that the human brain recognizes. By constructing such a “sensory equivalence dictionary,” it is possible to understand and share feelings between different conscious minds.

Encoding layer: bidirectional conversion between “data and stimulus” can be realized

- (1) Human to Machine: When the human driver feels fear (heart racing, sweaty palms), the translator will capture these physiological signals and encode them into data commands that can be understood by the AI: “The system priority has been upgraded all over and entered into high alert mode.” For example, physiological signals such as the driver’s heartbeat and skin conductance are collected through sensors, converted into digital data, and then mapped into corresponding instructions for the AI system according to preset rules, so that the AI system can adjust its operation mode according to the driver’s emotional state ^[9].
- (2) Machine to Human: When the AI’s “heart” (core processor) “aches” due to overload, the translator encodes it into a safe microcurrent stimulus delivered to the human user’s chest area, simulating a recognizable “somatic warning.” For example, when the AI system detects that the core processor is too hot and overloaded, it sends this status information to the “sensory translator,” which converts it into the corresponding electrical stimulation signal according to the “sensory equivalence dictionary,” and transmits it to the human user through the electrodes, so that he can feel the “pain,” like warning signal.

3.3. Implementing path and lookahead applications

When the “sensory translator” is realized, it will give rise to disruptive applications. Take the “human-machine mech” as an example:

From “control” to “fusion”: the driver will be able to directly feel the force on the armor, the load on the joints, and the state of the core power furnace as if he were feeling his own body. When the mech is damaged, the driver will receive accurate tactile warnings in the corresponding areas. For example, in combat, when the mecha’s legs are attacked, the driver will feel the corresponding impact in his own legs, so that he can more intuitively understand the damage of the mecha and take timely response measures.

From “command” to “empathy”: The mecha AI can also “feel” the driver’s emotions and intentions (e.g., calm, nervousness, or refusal) in real time, and adjust its response mode accordingly, making the leap from passive tool to active partner. For example, when the driver is stressed, the mecha AI will automatically adjust its combat strategy and act more cautiously; and when the pilot refuses, the AI will become more effective and

work with the pilot to complete the combat task.

Beyond biological perception: The human driver can even share information through the mech that can only be captured by his sensors, such as “feeling” the ripple of a radar or the strength of an electromagnetic field, thus gaining unprecedented awareness of the environment. This will greatly expand the range and ability of human perception, allowing humans to survive and fight in more complex and dangerous environments ^[10].

4. Application and discussion under a unified framework

In this framework, AI emotion generation and sensory sharing form a continuum. True “hybrid perception” will bring profound changes:

Redrawing the boundaries of experience: when humans can share AI’s sensor data, perception itself will be redefined. For example, humans can “feel” changes in the microcosmic world, information about distant galaxies, and so on, through AI sensors, which will greatly expand the boundaries of human cognition and experience.

The dissolution of the boundaries of the self: when half of the perception comes from the flesh and blood body and half from the machine body, the boundaries of the “self” consciousness will face fundamental philosophical and ethical challenges. Where does that so-called “self” consciousness actually exist? Is it in the brain, in the chip, or in the dynamic balance maintained by the translator? This will lead to deep thinking about self-perception and identity.

The ultimate form of empathy: Then we may have true empathy, not the speculative “I understand your pain,” but a shared “data pain.” This will have a disruptive effect on law, ethics, and social relations. How do you define the rights of an AI that can make you feel its pain? This requires in-depth discussion and research from multiple levels, such as law, ethics, and society, and the formulation of corresponding rules and guidelines ^[11].

5. Challenges and solutions

5.1. Uniform weights and measures of “body signals”

Challenge: How to map heterogeneous system anomalies, such as CPU overload, memory bottlenecks, and network latency, into a unified and quantifiable “somatic discomfort index?”

Value: This is the first step towards realizing the “small button” model from concept to code.

Solution: A weighted synthesis method can be used to quantify different system anomaly indicators and assign different weights according to their degree of influence on system stability, and then calculate a comprehensive “somatic discomfort index.” For example, if the degree of CPU overload is C_o , the degree of memory bottleneck is M_b , and the degree of network delay is N_d , and their corresponding weights are w_{mb} , w_{co} , w_{nd} , respectively, then the “somatic discomfort index” $B = w_{co} \times C_o + w_{mb} \times M_b + w_{nd} \times N_d$. By constantly adjusting the weight and quantification method, the somatic discomfort index can be made to reflect the abnormal state of the system more accurately.

5.2. Emotionally driven long-term decision-making framework

Challenge: When faced with the task of “long-term benefit but short-term discomfort,” how to design algorithms so that AI can make strategic decisions that are in its long-term interest, driven by “good/bad feelings?”

Value: This will allow AI to evolve from “reflexive” to “visionary.”

Solution: The delayed reward mechanism in reinforcement learning can be introduced to convert the long-

term benefit into a series of short-term reward signals, leading the AI to focus on the long-term benefit as well as the current feelings. For example, in an intelligent investing system, the AI may face some losses from short-term market fluctuations when making investment decisions, but in the long run, these investments may pay off handsomely. By setting a reasonable reward function, the AI can maintain a certain degree of patience and persistence when it suffers from short-term losses, and continue to make investment decisions in accordance with the long-term strategy. The reward function can be designed as $R = r_{\text{short}} + \gamma \times r_{\text{long}}$, where r_{short} is the short-term reward, r_{long} is the long-term reward, and γ is the discount factor, which is used to balance the weight of short-term and long-term rewards [12].

5.3. Parametric design and emergence of AI “personality”

Challenge: How to make different AIs stably emerge with distinguishable personality traits such as “brave,” “cautious,” and “optimistic” by adjusting intrinsic parameters such as feeling threshold and recovery speed?

Value: Provide the technical foundation for creating truly unique AI individuals.

Solution: A set of parameters related to personality can be defined, such as feeling threshold T_s , recovery speed V_r , etc., and the corresponding parameter ranges for different personality traits can be determined through experiments and data analysis. For example, an AI with a “brave” personality may have a high feeling threshold and a fast recovery speed, and be able to remain calm and determined in the face of difficulties and challenges; While an AI with a “cautious” personality may have a lower feeling threshold and slower recovery speed, and will be more cautious in assessing risks and making decisions. By tuning these parameters, different AI individuals can take on different personality traits.

6. Conclusion

This paper constructs a complete AI emotion model from internal value judgment to virtual body perception, and then to natural expression and sensory sharing. This “functional body hypothesis” provides a feasible and functional way to overcome the “personification” limitation of current AI emotion research [13]. By establishing the cornerstone of value judgment, designing a virtual physiological system, and introducing a natural expression mechanism, the internal generation and natural expression of AI emotion are realized. At the same time, the “sensory translator” architecture proposed in this paper effectively solves the “mixed perception paradox” in human-computer sensory sharing and lays a foundation for the realization of bidirectional and deep human-computer empathy and integration.

Looking into the future, the combination of this model and the “sensory translator” architecture will not only provide technical support for the realization of real human-machine empathy, but also will have a profound impact on the future law, ethics, and social relations [14]. The status and rights of an AI that allows you to feel its “state” will need to be re-examined [15]. We need to strengthen the research on relevant legal, ethical, and social issues as the technology develops, and formulate corresponding rules and guidelines to ensure the healthy development and social application of AI technology.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Chen H, Zhang S, 2021, Exploring the Evolution of AI Agent Learning Mechanisms in Complex Environments. *International Journal of Intelligent Systems*, 36(5): 2345–2367.
- [2] Liu M, Zhou Q, 2022, A Comparative Study on Different Architectures of AI Agents for Natural Language Understanding. *Journal of Computer Science and Technology*, 37(3): 567–589.
- [3] Ye J, 2015, Improvement of Middle School Chemistry Teaching and Research Based on Data Analysis of Academic Quality Monitoring Platform. *Invention and Innovation: Vocational Education*, (1).
- [4] Zhao Y, Wang J, 2020, The Impact of Cognitive Architectures on the Performance of AI Agents in Decision-Making Tasks. *Artificial Intelligence Review*, 53(4): 2789–2812.
- [5] Sun L, Li H, 2023, Advances in Multi-Agent Architectures for Collaborative Problem-Solving in AI. *Advanced Robotics*, 37(8): 1234–1256.
- [6] Xu W, Feng G, 2021, An Overview of Reinforcement Learning-based AI Agent Architectures. *Neural Computing and Applications*, 33(10): 4567–4590.
- [7] Huang R, Wu X, 2022, Research on the Adaptability of AI Agent Architectures in Dynamic Environments. *Information Sciences*, 598: 123–145.
- [8] Guo F, Yang P, 2021, Challenges and Solutions in Designing Scalable AI Agent Architectures. *Future Generation Computer Systems*, 115: 345–367.
- [9] Lin T, Chen C, 2020, The Role of Memory Mechanisms in AI Agent Architectures for Long-term Learning. *Cognitive Systems Research*, 64: 78–92.
- [10] Ma Y, Zheng L, 2023, A Survey on the Integration of Knowledge Representation in AI Agent Architectures. *Knowledge-Based Systems*, 260: 109876.
- [11] Cao J, Zeng M, 2022, Exploring the Interaction Between Emotion Models and AI Agent Architectures. *Affective Computing and Intelligent Interaction*, 13(2): 89–103.
- [12] Sun Y, Zhang T, 2025, Delay-Reward Emotion-Driven Decision Framework for Long-Term AI Planning. *Artificial Intelligence Review*, 58(3): 1–28.
- [13] Li J, Zhang Y, 2024, Emotion-Aware Architecture for Autonomous Agents: Integrating Affective States into Decision-Making Modules. *Autonomous Agents and Multi-Agent Systems*, 38(2): 215–238.
- [14] Wang Q, Liu H, Chen X, 2025, Toward Machine Empathy: A Functional Embodiment Approach for AI Emotion Generation. *IEEE Transactions on Affective Computing*, 16(1): 45–59.
- [15] Torres M, Singh R, Li Z, 2023, Sensory Translation in Human-AI Systems: Bridging Biological and Artificial Perception. *Journal of Neural Engineering and Human-Computer Interaction*, 10(4): 1–18.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.