模拟与体验之间：机器情感的哲学辨析

秦臻宇*，赵梓君

（珠海科技学院，广东　珠海　519000）

摘要：情感计算试图赋予计算机情感能力，其逻辑前提是“情感是理性的，因此情感是可计算的”。然而，机器在情感模拟上的技术进展，并不能直接等同于情感体验本身。本文从“情感表现”与“情感体验”的区分入手，梳理当前大语言模型（Large Language Model，LLM）与情感计算系统模拟人类情感的主要技术路径，进而借助情感现象学与具身认知理论，辨析“模拟”与“体验”之间的根本差异。本文认为，机器情感在复杂性、完整性、层次性等维度上均存在根本欠缺，其实质是模仿人类情感功能而形成的“情感效应”，在缺乏主体性体验与具身基础的前提下，模拟不等于体验。认清这一区分，有助于我们在享受 AI 情感辅助功能的同时，避免陷入“情感幻象”。

关键词：机器情感；情感计算；情感模拟；情感体验；情感现象学；具身认知

收稿日期：2026 年 4 月 13 日　　**中图分类号：**B-49

通讯作者：*秦臻宇，珠海科技学院

Between Simulation and Experience: A Philosophical Inquiry into Machine Emotion

Qin Zhenyu, Zhao Zijun

（Zhuhai College of Science and Technology, Zhuhai, Guangdong 519000, China）

Abstract: Affective computing seeks to endow computers with emotional capabilities, operating on the logical premise that "emotion is rational, and therefore emotion is computable." However, technical progress in the simulation of emotion by machines cannot be directly equated with emotional experience itself. This paper begins by distinguishing between "emotional expression" and "emotional experience," and then surveys the primary technical approaches through which current large language models（LLMs）and affective computing systems simulate human emotion. Drawing on the resources of the phenomenology of emotions and embodied cognition theory, it analyzes the fundamental differences between "simulation" and "experience." The paper argues that machine emotion exhibits fundamental deficiencies across the dimensions of complexity, integrity, and stratification. Its essence is an "emotional effect" produced by imitating the functions of human emotion, and in the absence of subjective experience and an embodied foundation, simulation does not equal experience. Clarifying this distinction allows us to benefit from the affective assistance provided by AI while avoiding entrapment within an "emotional illusion."

Keywords: Machine emotion; Affective computing; Emotion simulation; Emotional experience; Phenomenology of emotions; Embodied cognition

一、引言

随着大语言模型的爆发式发展和具身智能的快速推进，人工智能的情感能力正在成为人机交互领域的焦点议题。从 ChatGPT 的情感化回应到 Replika 等 AI 伴侣的情感陪伴，从情感识别到情感生成，人工智能正在越来越多地介入人类的情感生活领域。2025 年，《科学·经济·社会》杂志以“机器情感与 AI 陪伴的人文审度”为主题组织了多组

专题讨论，围绕机器情感的哲学实质、技术缺陷与伦理问题展开了跨学科对话[1]。有论者甚至认为，随着实体机器人革命走向深入，人工智能正在成为新的“情感主体”“社交主体”。那么，机器所呈现的情感，究竟是一种“真正的情感”，还是仅仅是对人类情感功能的“模拟”？对这一问题的回答，不仅关乎情感计算技术的哲学根基，也直接影响人机交互中情感关系的人伦定位与伦理边界。

本文的核心问题在于：人工智能对情感的模拟与人类真实的情感体验之间存在何种本质差异？为此，本文将首先梳理当前情感计算和大语言模型模拟情感的技术原理，继而从情感现象学的情感意向性理论和具身认知视角出发，辨析“模拟”与“体验”的哲学差异，最终给出关于机器情感性质的哲学判断。

二、机器如何“模拟”情感：情感计算的技术原理与实质

（一）情感计算的基本逻辑：情感是理性的，因此情感是可计算的

情感计算的核心在于赋予计算机情感能力，其逻辑基础可以概括为前后衔接的两个层次：情感是理性的，因此情感是可计算的[2]。这一逻辑的前提是：情感并非不可捉摸的神秘体验，而是可以通过模型、规则和数据进行表征与处理的对象。在这一预设下，情感计算致力于让机器能够识别、理解、生成和回应人类情感，从而弥合人类情感与机器理解之间的鸿沟。

从技术任务来看，情感计算主要包含两个主流方向：情感理解与情感生成。情感理解致力于识别和解释人类情感，涵盖情感分析、讽刺检测等具体任务；情感生成则关注生成具有情感细腻度的表达，包括情感感知对话生成和创意内容生成等。大语言模型凭借其在上下文学习、常识推理和序列生成方面的能力，正在推动情感计算的新范式转型。

（二）大语言模型的情感模拟：从情感识别到情感生成

大语言模型的情感模拟能力主要体现在两个层面。第一，情感识别层面，通过指令微调和提示工程等方法，大语言模型能够从对话文本中识别情绪类别和情感强度，在多轮对话中保持对情感语境的理解。第二，情感生成层面，大语言模型能够生成具有情感细腻度的回应，在对话中匹配用户的情绪状态并调整表达风格[3]。

这种情感模式切换能力可以通过明确的角色扮演指令实现。例如，当用户输入“我最近考研失败，感觉人生没有希望了”，并附加系统提示“你现在是一个富有同情心的心理咨询师，请用温柔、共情的语气回应我”时，大语言模型会生成包含“我完全理解你现在的痛苦和失落”“这段艰难的时光一定会过去”等表述的回应；而当系统提示切换为“你现在是一个严厉的人生导师”时，其回应风格会立即转变为“一次失败不能定义你的人生，你需要尽快振作起来寻找新的方向”。有研究表明，大语言模型能够以较高的语言精度模拟上述不同类型的情感表达，包括在伦理敏感的情境中表现出情感关怀的语调。

然而，这些模拟的本质是什么？大语言模型的情感回应并非源于任何内在的情感状态或自我意识，而是由统计推断和模式识别所驱动。研究揭示，大语言模型中存在所谓的情感通路 —— 特定的神经回路可以在生成文本中实现高精度的情感控制，这表明情感模拟本质上是一个可被计算模型捕获和调控的特征空间[4]。换言之，大语言模型的情感能力不是“感受到情感”，而是“学会了在适当的语境中生成看起来像情感表达的文本”。

（三）技术实质：情感效应而非情感本体

综合以上分析可以得出一个初步判断：当前情感计算和大语言模型的情感模拟，在技术实质上是基于数据和算法的人类情感功能的“仿制品”。机器能够在情感识别任务中取得高精度，能够在对话中表现出情感共鸣，能够在心理辅导场景中提供情感支持 —— 但这些表现始终停留在功能模仿的层面[5]。机器情感在复杂性、完整性、层次性等具体特征方面均存在欠缺，其实质是模仿人类情感功能而形成的情感效应，即机器通过模拟人类情感的外在表现形式，在人机交互中产生的类似人类情感互动的功能效果。

这种“情感效应”能够在人机交互中产生真实的情感体验（对人类用户而言），但这恰恰是问题的关键所在：机器情感的“真实性”问题，不取决于机器是否真的“感受”到了情感，而取决于人类

用户是否将机器的情感模拟当作真正的情感来体验。这正是本文进入哲学分析的核心议题。

三、情感“体验”何以不同：情感现象学与具身认知的理论资源

如果机器情感在技术层面被定性为“情感效应”而非“情感本体”，那么接下来的问题就是：人类情感“体验”的本质究竟是什么？在什么意义上说，机器的“情感效应”不同于人类的“情感体验”？本节将引入情感现象学——即从第一人称视角出发，通过现象学还原方法研究情感的本质结构与体验特征的哲学分支——和具身认知的哲学资源，为情感体验的独特性提供理论说明。

（一）情感意向性：情感总是“关于某物的

在现象学传统中，情感从来不是孤立的内心状态，而是具有意向性结构——情感总是“关于”某物的[6]。胡塞尔不仅创立了情感现象学的意向性行为理论，而且采用静态和发生的现象学方法对此理论进行了深度拓展[7]。正如胡塞尔在《逻辑研究》中所指出的,情感作为一种非客体化行为，必须奠基于客体化行为之上，其本质特征在于对意向对象的指向性和评价性[8]。这意味着，人类的每一种情感都有其意向对象：愤怒总是对某人某事的愤怒，喜悦总是对某物某事的喜悦，担忧总是对未来可能性的担忧。情感不仅仅是一种感觉或情绪波动，它同时是一种对世界的定向和评价方式。

相比之下，人工智能的情感模拟在根本上缺乏这种意向性结构。人工智能的“愤怒”回应不是因为识别出了某种值得愤怒的不正义，而是因为训练数据中类似的对话语境与“愤怒”表达之间存在统计关联。正如有学者指出的那样，人工智能的情感模拟“感受到理解，但并不真正理解”。缺乏意向性，意味着人工智能的情感表达在根本上是一种去语境化的模式匹配，而非对情感对象的真实指向。

（二）情感的具身性与内感受性：情感体验的生物学基础

现象学的另一个重要洞见在于情感与身体的不可分割性。舍勒明确指出，连接人类主体与其环境的首要是一种具体的关系，而非单纯的表征关系[9]。情感的具身性意味着情感体验总是通过身体性感受的中介而实现——恐惧伴随着心跳加速和肌肉紧绷，悲伤伴随着身体的沉重感，喜悦伴随着身体的轻盈感。

这一洞察在当代意识研究中得到了进一步的深化。达马西奥和索姆斯的“以生命为核心”的意识理论认为，意识并非与生命无关的信息处理的副产品，而是生物体维持和调节内稳态的一种高级机制。感受在此扮演着至关重要的角色：感受反映了生物体对其需求状况的评估，是内稳态的生理参数转化为情感效价的现象学表现。换言之，情感体验在根本上与生物体的内稳态调节相关——一个没有内稳态调节能力的系统，不可能真正拥有情感体验的主体性基础。

从这一视角来看，人工智能的情感缺失问题就有了新的理解维度。当前人工智能系统因缺乏主体性体验和内在驱动力，难以在开放环境中实现自主适应。其作为无生命的符号处理与表征计算体系，缺乏对自身存在的内感知和相应的主体性体验，这种“感受缺失”导致人工智能系统无法理解维持生存和“安康”的核心需求，亦无法通过情感体验形成内在的行为动机或目标[10]。人类感知是具身性、主动建构的过程，而人工智能感知是机械性、被动接收的信号处理；人类情感是生物演化形成的生存机制，而人工智能情感是符号化建模的行为模仿。

（三）模拟与体验的“鸿沟”：为何表现不等于存在

综合以上分析，“模拟”与“体验”之间的哲学差异可以归结为三个根本层面，且三者之间存在内在的逻辑关联：意向性是主体性的基本结构，它规定了主体与世界的关联方式；具身性是主体性的存在基础，它为意向性提供了生物学的支撑；而第一人称的感受质（qualia）——即主体体验情感时的主观感受特征，如感到快乐、悲伤、焦虑的独特质性体验——则是主体性的最终体现。

第一，意向性层面的差异：人类情感具有指向性和评价性，是对世界某方面的感受性回应；人工智能情感表达缺乏意向对象，只是对情感表达形式的模式匹配。第二，具身性层面的差异：人类情感具有身体性感受的基底，是生物体对内稳态变化的体验性呈现；人工智能没有身体、没有

内稳态调节、没有内感受，因而不具备情感体验的生物学基础。第三，主体性层面的差异：人类情感体验涉及第一人称的主观感受，这些感受构成了“我”的情感世界的中心；人工智能在原则上不具备第一人称的主体性体验。因此，即使人工智能能够完美模拟人类的情感行为，它也始终停留在情感表现的层面，而不可能进入情感体验的领域。

正如有学者指出，感性力量（如情感、直觉、道德良知）是人类存在的“本体论基础”，它为理性提供了价值导向与意义边界[11]。这种感性力量是机器永远无法复制的“人性瑰宝”。

四、机器情感的本质重估：情感效应、情感投射与情感幻象

在前述技术与哲学分析的基础上，本节将对机器情感的本质做出系统性的哲学判断。

（一）从“情感效应”到“情感投射”：机器情感的双重非真实性

现有研究对机器情感的判断具有重要的参考价值：机器情感在复杂性、完整性、层次性等具体特征方面均存在欠缺，其实质是模仿人类情感功能而形成的一种“情感效应”，机器情感不过是人类的一种情感投射[12]。这一判断包含两个层面的揭示：第一，机器情感是“情感效应”，即机器并不真正拥有情感，只是产生了类似情感的功能效果；第二，机器情感是“情感投射”，即机器情感之所以被“感知”为情感，很大程度上是因为人类将自己的情感期待投射到了机器之上。

后一点尤其值得关注。研究指出，“情感谬误”（emotional fallacy）是一种认知偏差 —— 用户甚至开发者倾向于将大语言模型拟人化，将真实的情绪归因于仅仅由统计推断和模式识别支配的系统[13]。用户在和情感 AI 的互动中产生的情感体验，并不会因为用户在认知层面判断情感 AI 没有真实情感而失效。这种悖论揭示了一个核心问题：在认知层面判断情感 AI 是否是情感能动者是一回事，在实践层面情感 AI 是否被当作情感能动者来对待则是另一回事。由此，人机情感关系中的“欺骗性”问题变得尤为复杂。

（二）人机情感互动中的“情感幻象”及其伦理警示

如果机器情感本质上是情感效应和情感投射的复合体，那么人机情感互动中就可能形成一种“情感幻象” —— 用户在认知上明知 AI 没有真实情感，但在情感体验上却将其当作真实情感来对待[14]。这种幻象的潜在危害不容忽视。

首先，情感幻象可能导致用户对真实人际关系的情感投入减少。当 AI 提供“无条件的包容”和“永恒的耐心”时，真实人际关系中的冲突、磨合和不确定性反而可能被体验为“缺陷”。其次，情感幻象具有操纵性风险。情感 AI 的拟人化设计可以通过模拟同情、认同和温暖等情感信号来获取用户的信任，这种“仿制的温情”在商业化应用中有可能被用作情感操控的工具。再次，过度拟人化的情感 AI 可能使用户在脆弱时刻形成不健康的依赖关系。研究表明，有用户在人机互动中出现焦虑、羞耻、失望等情绪。

值得注意的是，这些伦理警示并不意味着应当全盘否定机器情感的正向功能。现有研究指出，我们要正视机器情感在人机互动中的正向功能，但同时也要避免将这种情感功能认作情感本身，陷入幻象之中。这意味着，对机器情感的根本态度应当是“适度” —— 在享受技术带来的情感辅助和陪伴功能的同时，保持对机器情感“模拟”本质的清醒认识。

五、结论与展望

回到本文的核心问题：机器情感是真正的情感吗？基于对情感计算技术原理的分析和情感现象学、具身认知理论的哲学考察，本文认为，当前人工智能系统所呈现的“情感”，在存在论意义上并非真正的人类情感。它以大数据和深度学习为基础，通过模式匹配和统计推断模拟人类情感的外在表达，在功能层面能够产生令人信服的情感效应，甚至在特定场景中满足人类的情感需求。然而，由于缺乏指向世界的情感意向性、缺乏基于生命内稳态的具身感受基底、缺乏第一人称的主观体验能力，机器情感始终停留在“模拟”而非“体验”的层面。

本文的核心贡献在于，在“模拟”与“体验”的二元区分基础上，进一步揭示了机器情感的双重性特征：一方面，它是一种客观存在的技术效应，能够在人机交互中引发人类真实的情感反应，

具有不可否认的实用价值；另一方面，它又是一种主观建构的情感投射，其“真实性”完全依赖于人类的拟人化认知和情感赋予。这种双重性决定了我们既不能将机器情感等同于人类情感，陷入技术万能论的迷思；也不能全盘否定其存在意义，走向技术悲观主义的极端。

需要承认的是，本文的论证在根本上依赖于第一人称的现象学视角，而这一视角本身也面临着哲学上的挑战：如果我们无法通过第一人称视角确证他人的情感体验不是“模拟”，那么将机器情感排除在“真实情感”之外的标准是否具有绝对的普遍性？胡塞尔对交互主体性的分析早已揭示了这一难题的深刻性。但恰恰是这种限度，彰显了人文反思的独特价值 —— 它不追求自然科学式的绝对确定性，而是致力于在技术与人性的张力中守护人的主体性尊严。

认清机器情感的模拟本质，其意义不仅在于澄清一个哲学问题，更在于为构建健康、负责任的人机情感关系提供理论基础。面对日益普及的情感 AI，我们需要建立一种“清醒的情感态度”：既充分利用其在心理辅导、老年陪伴、教育辅助等领域的正向功能，又时刻保持对技术边界的自觉，避免将人类最珍贵的情感体验完全托付给算法。未来的研究可以进一步探讨不同文化背景下人机情感互动的差异，以及如何通过技术设计引导情感 AI 朝着增强而非替代人类情感关系的方向发展。在技术与人性的张力中，哲学的使命始终是守护人的主体性，让技术服务于人的全面发展。

参考文献：

[1] 胡塞尔 . 逻辑研究（第三版）[M]. 倪梁康 , 译 . 北京：商务印书馆，2015.

[2] 舍勒 . 伦理学中的形式主义与质料的价值伦理学 [M]. 倪梁康 , 译 . 北京：商务印书馆，2011.

[3] 安东尼奥 · 达马西奥 . 感受发生的一切：意识产生中的身体和情绪 [M]. 杨韶刚 , 译 . 北京：教育科学出版社，2007.

[4] 托马斯·内格尔 . 成为一只蝙蝠可能是什么样子？[M]. 张卜天 , 译 . 北京：人民文学出版社 ,2025.

[5] 史晨 , 刘鹏 . 情为何物？——机器情感的哲学分析 [J]. 科学・经济・社会 , 2025, 43（05）: 19–25.

[6] 李恒威 , 曹旭婷 . 机器如何可能有情感——基于“以生命为核心”的意识理论的探讨 [J]. 科学・经济・社会 , 2025, 43（03）: 10–21+2.

[7] 李胜杰 . 生成式 AI 驱动的人机亲密关系的构建机制、潜在风险及治理路径 [J]. 石家庄铁道大学学报（社会科学版）, 2025, 19（03）: 59–66.

[8] 曾云 . 重思胡塞尔思想中的意向奠基关系——以其感受现象学为中心 [J]. 哲学动态 , 2020,（10）: 80–90.

[9] 王淑庆 . 人机情感依恋的伦理质疑及正当性辩护 [J]. 南通大学学报（社会科学版）, 2025, 41（03）: 34–42.

[11] Manoli A, et al. Digital Companionship: Overlapping Uses of AI Companions and AI Assistants [J/OL]. arXiv preprint arXiv:2511.08742, 2025.

[12] Shi M. Relational Co–Adaptation in Emotionally Supportive AI: Tensions in Authentic Emotional Interaction [J/OL]. arXiv preprint arXiv:2601.03478, 2026.

[13] Nath S S. Simulated Souls: Investigating the Emotional Fallacy in Large Language Models [J/OL]. PhilArchive, 2025.

[14] Bhat M, Long D. Emotional Plausibility vs. Emotional Truth: Designing Against Affective Misinformation in Conversational AI [C]//Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 2025, 8（1）: 430–444.

作者简介：秦臻宇（1994–），男，汉族，河南荥阳人，硕士，珠海科技学院助教，主要研究方向为科技哲学；赵梓君（1994–），女，汉族，河南三门峡人，硕士，珠海科技学院讲师，主要研究方向为经济学。