Correspondence

https://doi.org/10.1038/s41562-026-02558-6

Large language models do not have emotions

Check for updates

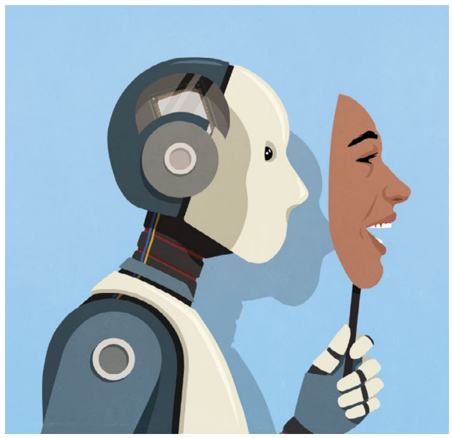
Do large language models (LLMs) have emotions? Could they ever have emotions? These questions captivate because they cut to the heart of our deepest fascination with what makes us human, and they deserve a serious answer.

One response is to say that LLMs do not (and could not ever) have emotions, because LLMs cannot feel[1,2]. If feeling is a necessary condition for the occurrence of emotions, the idea that LLM have (or will ever have) emotions could be dismissed. However, we and many other affective scientists do not see conscious feelings as a necessary condition of emotions in humans or other animals. We therefore do not regard the presence of feelings as an appropriate criterion for whether an LLM has emotions.

What might be more appropriate criteria? To answer this question, we adopt a functional perspective[2–4], asking what functions emotions serve and whether an LLM has processes that serve those functions. We argue that emotions serve two central functions. Specifically, emotions shape how we (1) interpret and (2) respond to the world.

Starting with the first function, emotions help us to make sense of the world via appraisal, a process by which the individual evaluates a situation in ways that determine which emotion (if any) is generated in response to that situation. Different appraisal profiles – for example, based on whether a situation is seen as desirable, caused by oneself or another, and manageable – give rise to different emotions[5,6]. In this respect, there is a point of convergence between LLM processing and human emotion. A non-peer-reviewed preprint by Anthropic has recently claimed that LLMs appear to evaluate the contextual features of a conversation and generate internal representations that track emotional significance[4].

However, a closer look at these findings reveals a key point of difference with the human case. A central insight from research on emotions is that emotions are context dependent. This means that similar situations can at different times give rise to different appraisal profiles. The same situation can generate fear in one person and excitement in another, depending on how it is appraised. So, to interpret the discovery of distinct,



consistent internal representations for different emotion concepts as evidence of distinct emotional states implies that each emotion has a reliable neural signature. But this misses the flexibility that is central to emotion. The neuroscience literature reinforces this point. Despite repeated efforts, researchers have not found consistent neural signatures for discrete emotion categories[7]. Emotion representations in the brain are found to be distributed, context dependent and variable across individuals and contexts[8,9]. Variability, rather than consistency, is the norm.

The second function of emotions is that they shape how we respond to the world. Consider the process of fear: attention narrows, decisions accelerate and the body prepares for action. Emotions constitute a flexibly crafted package of changes that reorganizes how cognitive resources are allocated across multiple systems simultaneously. Emotions also orient the organism towards action. Anger motivates approach and confrontation, and fear motivates avoidance or flight. These action tendencies represent a genuine shift in motivational state, and change the probability, speed and character of actions in ways that no single cognitive representation would predict.

When thinking about whether LLMs have emotions, it is easy to focus on what the model produces in response to emotionally relevant situations. But measuring the content of outputs captures only the most visible aspect of an emotional response. What it misses is what emotions do to the processing that generates those outputs. Unlike humans, LLMs do not have a processing architecture that is dynamically reconfigured in response to emotional representations across time and situations. Each forward pass through the network is computationally fixed. This is importantly different from the way emotion reorganizes processing in biological systems, and to argue that LLM has processes that function like emotions conflates the outputs of emotional processing with the process itself.

What evidence would make a more compelling case? We suggest that what is needed is evidence that LLM emotions involve functionally interpretable changes in processing: in attention, processing speed, and response type and likelihood. Crucially, this criterion does not require emotions to take any particular physical form. Consider collective emotions, which are emotions experienced by a group rather than a single individual. When a group faces a collective threat, the resulting fear-like state is not instantiated in any single body but distributed across a network of individuals. Yet it has genuine emotional properties: attention narrows towards the threat, decision-making accelerates and processing is reorganized around the emotionally relevant stimulus across the episode[10]. The substrate is social and distributed, but the functional signature is present.

A similar logic could apply to an LLM. In addition to showing situational and agent-related variability in reappraisal-like processes, one would need to show that appraisal-related emotion-like representations produce sustained, multisystem reorganization of processing, and not merely a shift in the probability of the next token. Perhaps one day LLMs will have these two core functional features of emotion. Until then, we are not ready to say that they have emotions.

Amit Goldenberg[1,2,3] & **James J. Gross**[4]

[1]Harvard Business School, Boston, MA, USA. [2]Department of Psychology, Harvard University, Cambridge, MA, USA. [3]Harvard Business School AI Institute, Boston, MA, USA. [4]Department of Psychology, Stanford University, Stanford, CA, USA.

e-mail: agoldenberg@hbs.edu

Published online: 24 August 2026

References

1. James, W. *Mind* **9**, 188–205 (1884).
2. Adolphs, R. & Andler, D. *Emot. Rev.* **10**, 191–201 (2018).
3. Keltner, D. & Gross, J. J. *Cogn. Emot.* **13**, 467–480 (1999).
4. Sofroniew, N. et al. Preprint at *transformer-circuits.pub* https://transformer-circuits.pub/2026/emotions/index.html (2026).
5. Uusberg, A., Taxer, J. L., Yih, J., Uusberg, H. & Gross, J. J. *Emot. Rev.* **11**, 267–282 (2019).
6. Ellsworth, P. C. & Scherer, K. R. in *Handbook of Affective Sciences* (eds Davidson, R. J. et al.) 572–595 (Oxford Univ. Press, 2003).
7. Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E. & Barrett, L. F. *Behav. Brain Sci.* **35**, 121–143 (2012).
8. Lee, K. M. & Satpute, A. B. *Soc. Cogn. Affect. Neurosci.* **19**, nsae043 (2024).
9. Satpute, A. B. & Lindquist, K. A. *Trends Cogn. Sci.* **23**, 851–864 (2019).
10. Goldenberg, A. *Perspect. Psychol. Sci.* **19**, 489–502 (2024).

Competing interests

The authors declare no competing interests.