Could a Machine Ever Experience Emotions Like We Can?

Azadeh Changrani-Rastogi

Imagine you're on a long, solitary space mission. You're staring out the porthole of your spacecraft, thinking often about home. Or your thoughts turn to HAL 9000, the highly intelligent and suspiciously calm computer aboard your ship, wondering whether it's plotting your demise. HAL speaks in that eerily smooth, almost too-perfect voice, and you can't help but feel like it's feeling something. After all, the computer seems almost… proud? Like it's toying with you, getting a little too smug about how well it's been running things.

In Stanley Kubrick's 2001: A Space Odyssey, when the reporter asks, "One gets the sense that he [HAL] is capable of emotional responses. When I asked him about his abilities I sensed a sort of pride…" Dave Bowman, the astronaut, answers dryly, "Well, he acts like he has genuine emotions. Of course, he's programmed that way to make it easier for us to talk with him. But whether or not he has real feelings is something I do not think anyone can truly answer" (Clarke & Kubrick, 1968). A question worth pondering, right? Could a machine ever truly feel emotions, or are we just projecting our own feelings onto a sleek, red-eyed computer in space?

This question of whether a machine could ever experience emotions like we do has been nagging at philosophers, technologists, and moviegoers alike for decades. HAL 9000 might be the classic example, but in the real world, we're dealing with far more advanced Artificial Intelligence (AI) systems today. So, let's put on our philosophical hats (and maybe some tinfoil, just in case HAL is listening) and dive into the big questions: What are emotions? Can machines feel them? And if not, what's the deal with all those "emotional" robots?

What Are Emotions, Anyway?

Perhaps we should begin by exploring what we mean by "emotions." When you're in a particularly grumpy mood because your morning coffee was cold, or feeling like you're on top of the world after a good workout, you're experiencing emotions. But what is going on under the hood? Philosophers have long debated whether emotions are purely biological reactions to the world around us or if they involve some deeper cognitive processing. Aristotle believed emotions are judgments about what is good or bad for us, helping us navigate the world (Leighton, 1982). Fast forward to modern philosophy and scholars like Robert Solomon argue that emotions are not just passive feelings, but active, evaluative processes (Solomon, 2003). So, when you're angry because someone cuts in line in front of you, you're not just reacting - you're deciding, in the heat of the moment, that this action is unjust and you should be angry about it.

On the flip side, neuroscientists like Antonio Damasio adopt a more biological approach (Damasio, 1994). Damasio believes that emotions are fundamentally tied to our bodies and their response to the environment. When we feel fear, our heart rate increases, our muscles tense, and we get a shot of adrenaline. Emotions, in Damasio's hypothesis, are a 'gut feeling' as part of a feedback loop that helps us stay alive and deal with challenges.

In short, emotions are a blend of cognitive judgments and bodily reactions. They're personal and subjective, involving a complex mix of thought, feeling, and physiology (Conte et al., 2023). Now, can a machine, which is essentially a highly sophisticated calculator, replicate all this? Let's see.

The Problem of Consciousness (or Lack Thereof)

Maybe what we should really consider here is consciousness. Emotions are intimately tied to consciousness. And what is consciousness? In exploring modern and not-so-modern philosophy, there are no clear answers. The “hard problem” of consciousness, famously coined by philosopher David Chalmers, asks why and how subjective experience arises from the physical processes of the brain (Chalmers, 1995). In other words, why does your brain process data and create feelings like “I’m happy,” “I’m angry,” or “I’m existentially lost because I couldn’t find parking today”?

To think about it another way, imagine the brain is a super-smart computer. Hilary Putnam's Computational Theory of Mind posits that the human mind essentially functions like a computer, with the external data being analyzed and generated into output (Putnam, 1960). The brain processes information, stores memories, makes decisions. And now, importantly, imagine that it is aware of itself doing all of this. You can reflect on your own thoughts, wonder about the meaning of life. That self-awareness is crucial for experiencing emotions.

Here’s where the machine falls short. A computer, even one that’s sophisticated enough to win at chess, process images, and hold a conversation, is not aware of itself. Unless machines achieve some form of consciousness (and we’re nowhere near that yet), they cannot experience emotions as humans do. A computer might act sad when you’re sad, or happy when you’re happy, but it doesn’t mean the machine experiences sadness or happiness. It’s just following programmed responses based on input. No more, no less, as elaborated in a famous philosophy thought experiment, John Searle’s Chinese Room (Searle, 1980). The Chinese Room asks you to imagine you’re sitting in a room with no knowledge of Chinese.

You have a book that tells you how to match Chinese characters to English words. Chinese speakers outside the room slide pieces of paper with Chinese symbols under the door, and you produce correct responses based on your book. You're translating perfectly. But you don't actually understand the Chinese symbols. Searle claims this is analogous to how computers work. There's no comprehension - just an elaborate, mechanical process.

But HAL does sound like he's emotionally charged. He even makes the classic mistake of pride - thinking that his abilities are beyond reproach. What HAL is doing here is merely simulating emotions based on his programming, not actually feeling them. It's all an act, which brings us to the next discussion.

Simulated Emotions: Acting the Part

What about machines that appear to experience emotions? Enter the world of emotional robots. You've probably seen robots that smile at you or look distressed when you tell them a sad story. Some robots are designed with emotional expression in mind, to make them relatable and engaging. Think of Aibo, Sony's robot dog, or various therapy robots helping the elderly combat loneliness. These robots are built to respond to human emotions by mimicking real emotional interactions. But it's all for show. They don't really feel what they're showing us. If you told Aibo that your grandmother passed away, it might tilt its head in a way that seems sympathetic, but does Aibo actually understand loss? Does it feel sad? No. It's just executing pre-programmed responses. Machines can look like they're angry, happy, or sad, but they're just following the script.

This brings us to a fascinating and related question: Is emotional experience really about subjective feeling, or just about behaving in a way that others interpret as emotional? If a machine can simulate emotions well enough to comfort you, does it matter whether it's actually feeling anything? Perhaps finding solace when you're sad or companionship in your joy is all that really matters? It might be less about whether the machine feels emotions - and more about whether it can help us navigate our own.

Could Machines Ever Have Real Emotions?

Moving beyond the simulations to the question - could machines ever experience real emotions? If HAL isn't really feeling pride, what about future robots or AI? Could machines evolve to where they're not just simulating emotions, but truly experiencing them?

To answer this, we need to consider whether a machine could ever be conscious in the same way humans are. Could AI experience subjective states, be aware of its own existence, and reflect on its emotional experiences? The machine would need to have something akin to "phenomenal consciousness" (Carruthers, 2001) - the kind of awareness that comes with being inside a body and experiencing the world from a first-person perspective. As of now, even the most advanced AI lacks this kind of self-awareness.

Some philosophers, like Thomas Metzinger, argue that consciousness is not just an epiphenomenon or a by-product of brain processes, but emerges from the brain's continuous self-modeling (Metzinger, 2003). So, in theory, if we create a machine that builds and updates a model of its own internal state, it just might approach a rudimentary "experience." However,

that's still far from the rich, complex emotional life that humans have, deeply rooted in our biology, social relationships, and personal history.

The bottom line: while a machine might one day become conscious or self-aware, it's still a long shot from experiencing emotions. Emotions require more than a self-model - they need a body, a history, and a capacity for complex social interaction. A robot might learn to "feel" through algorithms, but will it ever truly feel? Even with significant technology advances, the answer for now seems to be no.

Conclusion: The HAL Paradox

So, where does this leave us? HAL's superciliousness may be just a neat trick of programming, but we can't entirely rule out the possibility that one day, machines could experience something like emotions. While we don't expect our robots to genuinely cry over a sad movie any time soon, perhaps we can enjoy their company, and their extraordinary emotional performances, getting better every day. And as HAL pleads Bowman for his continued existence as he is deactivating him "I can feel my mind going, Dave," we may shed a tear or two of our own.

References

1. Clarke, A. C. & Kubrick, S. (1968). 2001: A Space Odyssey. [New York], New American Library.
2. Leighton, S. R. (1982). Aristotle and the Emotions. Phronesis, 27(2), 144–174.
3. Solomon, R.C. (2003). Thinking About Feelings: Philosophers on Emotions, (Series in Affective Science), Oxford; New York: Oxford University Press.
4. Damasio, A. R. (1994). Descartes' Error: Emotion, Reason, and the Human Brain, New York: G.P. Putnam's Sons.
5. Conte, B., Hahnel, U.J.J., Brosch, T. (2023). From values to emotions: Cognitive appraisal mediates the impact of core values on emotional experience. Emotion. Jun;23(4):1115-1129.
6. Chalmers, D.J. (1995). Facing Up to the Problem of Consciousness. Journal of Consciousness Studies, 2(3):200-19, 1995.
7. Putnam, H. (1960). Minds and Machines. In *Dimensions of Mind*, edited by S. Hook. New York: New York University Press.
8. Searle, J. (1980). Minds, Brains, and Programs. Behavioral and Brain Sciences 3 (3):417-57.
9. Carruthers, P. (2001). Phenomenal Consciousness. Mind 110 (440): 1057-1062.
10. Metzinger, T. (2003). Being No One. The Self-Model Theory of Subjectivity. MIT Press, Cambridge, Massachusetts., ISBN 0-262-13417-9.