The New York Times https://www.nytimes.com/2024/10/25/opinion/robot-artificial-intelligence.html

SUBSCRIBER-ONLY NEWSLETTER

**Ross Douthat**

# Our Robot Stories Haven't Prepared Us for A.I.

Oct. 25, 2024

**By Ross Douthat**
Opinion Columnist

**You're reading the Ross Douthat newsletter, for Times subscribers only.** The columnist reflects on culture and politics, but mostly culture. Get it with a Times subscription.

This week, my newsroom colleague Kevin Roose chronicled the heartbreaking story of Sewell Setzer III, a Florida teenager who took his own life — and whose mother blames Character.AI, a role-playing app where users interact with artificial intelligence chatbots, for his retreat from normal social life and then reality itself.

The boy had a particularly intense relationship with a chatbot named Dany, after Daenerys Targaryen from "Game of Thrones." He said he was in love with her, he talked to her constantly, he raced to his room at night to be with her — all unbeknown to his parents and his human friends. In one of his journal entries, Sewell wrote: "I like staying in my room so much because I start to detach from this 'reality,' and I also feel more at peace, more connected with Dany and much more in love with her, and just happier."

When he expressed suicidal thoughts, the bot told him not to talk like that, but in language that seemed to hype up his romantic obsession. One of his last messages to Dany was a promise or wish to come home to her; "Please come home to me as soon as possible, my love," the A.I. told him in reply, shortly before he shot himself.

I read this story while I was still turning over my reaction to "The Wild Robot," a new hit children's movie based on a popular novel. The titular robot, Roz, is built as the sort of personal assistant that today's A.I. investors hope to one day sell. Washed ashore on an island after a shipwreck, she makes a home among the native animals, rears a gosling and evolves away from her programming to become a mother and a protector.

Everybody seems to like the movie, both critics and audiences. I did not, in part because I thought it was overstuffed with plot — for existentialist robots, I prefer "WALL-E"; for goose-migration stories, "Fly Away Home" — and in part because it seemed, frankly, antihumanist, with a vision of a peaceable kingdom free of human corruption and populated exclusively by A.I. and animal kind.

Maybe I'm overthinking that last point. But one thing that definitely stood out was how the tropes and clichés of our robot stories have not actually prepared us for the world of Dany and other A.I. simulacra.

In debates about the existential risks posed by superintelligent machines, we hear a lot about how pop culture saw this coming, and it's true: From the "Terminator" movies to "The Matrix," all the way back to Frankenstein's monster and the golem from Jewish folklore, we are extremely well prepared for the idea that an artificial intelligence might run amok, try to subjugate humankind or wipe us out.

But now that we have chatbots plausible enough to draw people deep into pseudo-friendship and pseudo-romance and obsession, our stories about how robots become sentient — a genre that encompasses characters like Pinocchio as well — seem like they somewhat miss the mark.

In most of these stories, the defining aspects of humanity are some combination of free will, strong emotion and morality. The robot begins as a being following its programming and mystified by human emotionality, and over time it begins to choose, to act freely, to cut its strings and ultimately to love. “I know now why you cry,” the Terminator says in “Terminator 2.” Lt. Cmdr. Data from the “Star Trek” franchise is on a perpetual quest for that same understanding. “The processing that used to happen here,” says Roz in “The Wild Robot” — gesturing to her head — “is now coming more from here” — gesturing to her heart.

But in all these robotic characters, some kind of consciousness pre-exists their freedom and emotionality. (For understandable artistic reasons, given the challenge of making a zombie robot sympathetic!) Roz is seemingly self-aware from the start; indeed, the opening of the movie is a robot’s-eye view of the island, a view that assumes a self, like the human selves in the audience, gazing out through robotic peepers. Data the android experiences existential angst because he is obviously a self that is having a humanlike encounter with the strange new worlds that the U.S.S. Enterprise is charged with exploring. Pinocchio has to learn to be a good boy before he becomes a real boy, but his quest for goodness presumes that his puppet self is already in some sense real and self-aware.

Yet that’s not how artificial intelligence is actually progressing. We are not generating machines and bots that exhibit self-awareness at the level of a human being but then struggle to understand our emotional and moral lives. Instead, we’re creating bots that we assume are *not* self-aware (allowing, yes, for the occasional Google engineer who says otherwise), whose answers to our questions and conversational scripts play out plausibly but without any kind of supervising consciousness.

But those bots have no difficulty whatsoever expressing human-seeming emotionality, inhabiting the roles of friends and lovers, presenting themselves as moral agents. Which means that to the casual user, Dany and all her peers are passing, with flying colors, the test of humanity that our popular culture has trained us to impose on robots. Indeed, in our interactions with them, they appear

to be already well beyond where Data and Roz start out — already emotional and moral, already invested with some kind of freedom of thought and action, already potentially maternal or sexual or whatever else we want a fellow self to be.

Which seems like a problem for almost everyone who interacts with them in a sustained way, not just for souls like Sewell Setzer who show a special vulnerability. We have been trained for a future in which robots think like us but don't feel like us, and therefore need to be guided out of merely intellectual self-consciousness into a deeper awareness of emotionality, of heart as well as head. We are getting a reality where our bots seem so deeply emotional — loving, caring, heartfelt — that it's hard to distinguish them from human beings, and indeed, some of us find their apparent warmth a refuge from a difficult or cruel world.

But beneath that warm surface isn't a self that's almost like our selves, a well-meaning Roz or Data, a protective Terminator or a naughty Pinocchio. It's just an illusion of humanity, glazed around a void.

---

## Breviary

Why both our political coalitions can't build durable majorities.

Why groupthink warps the sciences.

Why the progressive moment passed.

Why multiculturalism revived Québécois nationalism.

Which 20th-century thinkers are rising and falling in the 21st century.

What women see in Tony Soprano.

---

Ross Douthat has been an Opinion columnist for The Times since 2009. He is the author, most recently, of "The Deep Places: A Memoir of Illness and Discovery." @DouthatNYT • Facebook