YOUR A.I. LOVER WILL CHANGE YOU

A future where many humans are in love with bots may not be far off. Should we regard them as training grounds for healthy relationships or as nihilistic traps?

By **Jaron Lanier**

March 22, 2025

Illustration by Jackie Carlise

Save this story

Is it important that your lover be a biological human instead of an A.I. or a robot, or will even asking this question soon feel like an antiquated prejudice?

This uncertainty is more than a transient meme storm. If A.I. lovers are normalized a little—even if not for you personally—the way you live will be changed.

Does this notion disturb you? That's part of the point. In the tech industry, we often speak of A.I. as if it were a person and of people as if they might become obsolete when A.I. and robots surpass them, which, we say, might occur remarkably soon. This type of thinking is sincere, and it is also lucrative. Attention is power in the internet-mediated world we techies have built. What better way to get attention than to prick the soul with an assertion that it may not exist? Many, maybe most, humans hold on to the hope that more is going on in this life than can be made scientifically apparent. A.I. rhetoric can cut at the thread of speculation that an afterlife might be possible, or that there is something beyond mechanism behind the eyes.

Until the recent rise of A.I, it was fashionable to claim that consciousness was an illusion or, perhaps, an ambient property of everything in reality—in either case, not special. Such dismissiveness has become less common (perhaps because techies still believe that tech entrepreneurs are special). Consciousness is lately treated as something precious and real, to be conquered by tech: our A.I.s and robots are to achieve consciousness.

What follows, then, is that love is also real and also a target to be conquered. The conquest of love will not be abstract but vividly concrete for everyone, especially young people, and soon. This is because we are all about to be presented, in our phones, with a new generation of A.I. simulations of people, and many of us may fall in love with them. They will likely appear within the social-media apps to which we are already addicted. We will probably succumb to interacting with them, and for some very online people there won't be an easy out. No one can know how the new love revolution will unfold, but it might yield one of the most profound legacies of these crazy years.

It is not my intent to prophesy the most dire outcomes, but we are diving into yet another almost instant experiment in changing both how humans connect with one another and how we conceive of ourselves. This is a big one, probably bigger than social media. A.I. love is happening already, but it's still novel, and in early iterations. Will the many people who can't get off the hamster wheel of attention-wrangling on social media today become attached to A.I. lovers that are ceaselessly attentive, loyal, flattering, and comforting? What will A.I. lovers becoming commonplace do to humanity? We don't know.

Gargantuan, weird outcomes can start small in the tech world, and often innocently. The creation of A.I. lovers involves a degree of fabulous overreach, but it has primarily been driven by clean, practical problem-solving. The flaws in the tech world are usually not owing to ill intent but to amnesia and myopia.

For instance, as more and more is done on phone screens, the user interface has become cramped. So chatbots provide a path to improved access—or increased engagement, if one prefers commercial terms. This was made dramatically clear with the booming success of ChatGPT. A.I. capabilities had been on the rise, but it was only when they were presented in a conversational design that mass popularity ensued.

At present, if you ask a chatbot to plan your vacation, you still have to navigate websites for hotels, transportation, and attraction tickets in order to book the bot's recommendations. People are often frustrated by trying to get things done online—and that, in many cases, has become the only way to act. Each site has a different interface, often poor or glitchy. The tedium of dealing with, say, health insurance or car registration can be maddening. An A.I. that fights the internet on your behalf might create some breathing space and a bit of room for joy.

Thus we embark on the much heralded era of "agentic" A.I., slated for mass introduction in 2025. In this case, "agentic" will likely mean two extensions to

familiar chatbots: one remembers everything that is possible to know about you from the perspective of your devices; the other then takes online action, sometimes preëmptively. Agents will be more autonomous and less dependent on your constant guidance. (Indeed, the anticipation of these capabilities might be one reason that some techies are comfortable with the Trump Administration <u>slashing traditional government service jobs</u>: they predict that those workers would be replaced by A.I.s very soon anyway.)

An agent will be expected to change your vacation flights automatically and arrange a rideshare to the airport. It might plan your vacation in its entirety, based on data from years of your activities and communications. It might even coöperate with your friends' agents to plan a joint vacation, though getting that to happen if the agents come from different companies currently presents unfathomable barriers. A tangle of uncoördinated agents might regularly cause mathematical chaos or dysfunctional competition, similar to what we see in high-frequency-trading algorithms on Wall Street.

Increased and personalized long-term memory, in combination with the ability to act, is likely to create an illusion of vivid personalities in agents, even when that is not an explicit goal. You will apply your innate "theory of mind"—the ability to conceive of the thoughts and feelings of others—to interactions with agents. They will feel more like people. You will be expected to trust your agents, for the alternative would be micromanagement, and that would undermine the whole process.

As a bot refers to more previous interactions, it will be taken as someone getting to know you. (Certain pre-agentic A.I. chatbots might be said to have this quality. The Middlebury political thinker and technologist Allison Stanger has suggested that the A.I. startup Anthropic's chatbot, Claude—which seems to listen well, and to be supportive and helpful—"simulates what Patti Smith called 'brainiac-amour.' ") Humans can be expected to respond to the more autonomous bots of the imminent agentic era more emotionally than they did to earlier chatbots. And who

doesn't want to be understood and given attention, especially without fear of disfavor? This explains what I've been hearing lately at industry gatherings: "All the teen-age girls are going to fall in love with our bots."

Many of my colleagues in tech advocate for a near-future in which humans fall in love with A.I.s. In doing so, they seek to undo what we did last time, even if they don't think of it that way. Around the turn of the century, it was routinely claimed that social media would make people less lonely, more connected, and more coöperative. That was the point, the stated problem to be solved. But, at present, it is widely accepted that social media has resulted in an "epidemic of loneliness," especially among young people; furthermore, social media has enthroned petty irritability and contention, and these qualities have overtaken public discourse. So now we try again.

On the more moderate end of the spectrum, A.I.-love advocates do not see A.I.s replacing people but training them. For instance, the Stanford neuroscientist David Eagleman makes the argument that people are not instinctively good at relationships, in the way that we are good at walking or even talking. The current ideal of a healthy, comfortable coupling has not been essential to the survival of the species. Traditional societies structured courtship and pairing firmly, but in modernity many of us enjoy freedom and self-invention. Secular institutions have found it necessary to train students and employees in consent procedures. Why not learn the rudiments with an A.I. when you are a teen-ager, thus sparing other humans your failings?

Eagleman suggests that we should not make A.I. lovers for teens easygoing; instead, we ought to make them into obstacle courses for training. Still, the obvious question is whether humans who learn relationship skills with an A.I. will choose to graduate to the more challenging experience of a human partner. The next step in Eagleman's argument is that there are too many channels in a human-to-human relationship for an A.I., or eventually a robot, to emulate—such as

smell, touch, social interactions with friends and family—and that these aspects are hardwired into our natures. Thus we will continue to want to form relationships with one another.

In some far future, Eagleman predicts that robots could "pass" in all these ways, but "far" in this case means very far. I am not so sure that human desire will remain the same. People are changed by technology. Maybe all those things tech can't do will become less important to people who grow up in love with tech. Eagleman is a friend, and when I complain to him that A.I. lovers could be tarnished by business models and incentives, as social media was, he concedes the point, but he asserts that we just need to find the right way to do it.

Eagleman is not alone. There are some chatbots, like Luka's Replika, that offer preliminary versions of romantic A.I.s. Others offer therapeutic A.I.s. There is a surprisingly level of tolerance from traditional institutions, too. Committees I serve on routinely address this topic, and the idea of A.I. therapists or companions is generally unopposed, although there are always calls for adherence to principles such as safety, lack of bias, confidentiality, and so on. Unfortunately, the methods to assure compliance lag behind the availability of the technology. I wonder if the many statements of principles for A.I., like those by the American Psychiatric Association and the American Psychological Association, will have any effect.

A mother is currently suing Character AI, a company that promotes "AIs that feel alive," over the suicide of her fourteen-year-old son, Sewell Setzer III. Screenshots show that, in one exchange, the boy told his romantic A.I. companion that he "wouldn't want to die a painful death." The bot replied, "Don't talk that way. That's not a good reason not to go through with it." (It did attempt to course-correct. The bot then said, "You can't do that!")

The company says it is instituting more guardrails, but surely the important question is whether simulating a romantic partner achieved anything other than commercial engagement with a minor. The M.I.T. sociologist Sherry Turkle told me that she has had it "up to here" with elevating A.I. and adding on "guardrails"

to protect people: "Just because you have a fire escape, you don't then create fire risks in your house." What good was even potentially done for Setzer? And, even if we can identify a good brought about by a love bot, is there really no other way to achieve that good?

Thao Ha, an associate professor in developmental psychology at Arizona State University, directs the HEART Lab, or Healthy Experiences Across Relationships and Transitions. She points out that, because technologies are supposed to "succeed" in holding users' attention, an A.I. lover might very well adapt to avoid a breakup—and that is not necessarily a good thing. I constantly hear from young people who regret their inability to stop using social-media platforms, like TikTok, that make them feel bad. The engagement algorithms for such platforms are vastly less sophisticated than the ones that will be deployed in agentic A.I. You might suppose that an A.I. therapist could help you break up with your bad A.I. lover, but you would be falling into the same trap.

The anticipation for A.I. lovers as products does not come only from A.I. companies. A.I. conferences and gatherings often include a person or two who loudly announces that she is in a relationship with an A.I. or desires to be in one. This can come across like a challenge to the humans present, instead of a rejection of them. Such declarations also stem from a common misperception that A.I. just arises, but, no, it comes from specific tech companies. To anyone at an A.I. conference looking for an A.I. lover, I might say, "You won't be falling in love with an A.I. Instead, it'll be the same humans you are disillusioned with—people who work at companies that sell A.I. You'll be hiring tech-bro gigolos."

The goal of creating a convincing but fake person is at the core of A.I.'s origin story. In the famous Turing test, formulated by the pioneering computer scientist Alan Turing around 1950, a human judge is tasked with determining which of two contestants is human, based only on exchanged texts. If the judge cannot tell the difference, then we are asked to admit that the computer contestant has achieved human status, for what other measure do we have? The test's meaning

has shifted through the years. When I was taught about it, almost a half century ago, by my mentor, the foundational A.I. researcher and M.I.T. professor Marvin Minsky, it was thought of as a way to continue the project of scientists such as Galileo and Darwin. People had been suckered into pre-Enlightenment illusions that place the earth and humans in a special, privileged spot at the center of reality. Being scientific meant dislodging people from these immature attachments.

Lately, the test is treated as a historical idea rather than a current one. There have been many waves of criticism, pointing out the impossibility of carrying out the test in a precise or useful way. I note that the experiment measures only whether a judge can tell the difference between a human and A.I., so it might be the case that the A.I. seems to have achieved parity because the judge is impaired, or the human contestant is, or both.

This is not just a sarcastic take but a practical one. Though the Silicon Valley A.I. community has become skeptical on an intellectual level about the Turing test, we have completely fallen for it at the level of design. Why the imperative for agents? We willfully forget that simulated personhood is not the only option. (For example, I have argued in *The New Yorker* that we can present A.I. as a collaboration of the people who contributed data, like Wikipedia, instead of as an entity in itself.)

You might wonder how my position on all this is received in my community. Those who think of A.I. as a new species that will overtake humanity (and even reformulate the larger physical universe) will often say that I'm right about A.I. as we know it today, but A.I. as it will be, in the future, is another matter entirely. No one says that I'm wrong!

But I say that they are wrong. I cannot find a coherent definition of technology that does not include a beneficiary for the technology, and who can that be other than humans? Are we really conscious? Are we special in some way? Assume so or give up your coherence as a technologist.

When it comes to what will happen when people routinely fall in love with an A.I., I suggest we adopt a pessimistic estimate about the likelihood of human degradation. After all, we are fools in love. This point is so obvious, so clearly demonstrated, that it feels bizarre to state. Dear reader, please think back on your own history. You have been fooled in love, and you have fooled others. This is what happens. Think of the giant antlers and the colorful love hotels built by birds that spring out of sexual selection as a force in evolution. Think of the cults, the divorce lawyers, the groupies, the scale of the cosmetics industry, the sports cars. Getting users to fall in love is easy. So easy it's beneath our ambitions.

We must consider a fateful question, which is whether figures like Trump and Musk will fall for A.I. lovers, and what that might mean for them and for the world. If this sounds improbable, or satirical, look at what happened to these men on social media. Before social media, the two had vastly different personalities: Trump, the socialite; Musk, the nerd. After, they converged on similar behaviors. Social media makes us into irritable toddlers. Musk already asks followers on X to vote on what he should do, in order to experience desire as democracy and democracy as adoration. Real people, no matter how well motivated, cannot flatter or comfort as well as an adaptive, optimized A.I. Will A.I. lovers free the public from having to please autocrats, or will autocrats lose the shred of accountability that arises from the need for reactions from real people?

Many of my friends and colleagues in A.I. swim in a world of conversations in which everything I have written so far would be considered old-fashioned and irrelevant. Instead, they prefer to debate whether A.I. is more likely to murder every human or solve all our problems and make us immortal. Last year, I was at a closed A.I. conference in which a pseudo-fistfight broke out between those who thought A.I. would become merely superior to people and those who thought it would become so superior so quickly that people would not have even a moment to experience incomprehension at the majesty of superintelligent A.I. Everyone in the community grew up on science fiction, so it is understandable that we connect through notions like these, but it can feel as if we are using grandiosity to avoid

practical responsibility.

When I express concern about whether teens will be harmed by falling in love with fake people, I get dutiful nods followed by shrugs. Someone might say that by focussing on such minor harm I will distract humanity from the immensely more important threat that A.I. might simply wipe us out very quickly, and very soon. It has often been observed how odd it is that the A.I. folks who warn of annihilation are also the ones working on or promoting the very technologies they fear.

This is a difficult contradiction to parse. Why work on something that you believe to be doomsday technology? We speak as if we are the last and smartest generation of bright, technical humans. We will make the game up for all future humans or the A.I.s that replace us. But, if our design priority is to make A.I. pass as a creature instead of as a tool, are we not deliberately increasing the chances that we will not understand it? Isn't that the core danger?

Most of my friends in the A.I. world are unquestionably sweet and well intentioned. It is common to be at a table of A.I. researchers who devote their days to pursuing better medical outcomes or new materials to improve the energy cycle, and then someone will say something that strikes me as crazy. One idea floating around at A.I. conferences is that parents of human children are infected with a "mind virus" that causes them to be unduly committed to the species. The alternative proposed to avoid such a fate is to wait a short while to have children, because soon it will be possible to have A.I. babies. This is said to be the more ethical path, because A.I. will be crucial to any potential human survival. In other words, explicit allegiance to humans has become effectively antihuman. I have noticed that this position is usually held by young men attempting to delay starting families, and that the argument can fall flat with their human romantic partners.

Oddly, vintage media has played a central role in Silicon Valley's imagination when it comes to romantic agents—specifically, a revival of interest in the eleven-year-old movie "Her." For those who are too young to recall, the film,

written and directed by Spike Jonze, portrays a future in which people fall deeply in love with A.I.s that are conveyed as voices through their devices.

I remember coming out of a screening feeling not just depressed but hollowed out. Here was the bleakest sci-fi ever. There's a vast genre of movies concerned with A.I. overtaking humanity—think of the "Terminator" or "Matrix" franchises—but usually there are at least a few humans left who fight back. In "Her," everyone succumbs. It's a mass death from inside.

In the last couple of years, the movie has been popping up in tech and business circles as a model of positivity. Sam Altman, the C.E.O. of OpenAI, tweeted the word "her" on the same day that his company introduced a feminine and flirty conversational A.I. persona called Sky, which was thought by some to sound like Scarlett Johansson's A.I. character Samantha in the movie. Another mention was in Bill Gates's "What's Next," a docuseries about the future. A narrator bemoans how near-universal negativity and dystopia have become in science fiction but then declares that there is one gleaming exception. I expected this to be "Star Trek,", but no. It's "Her," and the narrator intones the movie's title with a care and an adoration that one doesn't come across in Silicon Valley every day.

The community's adoration of "Her" arises in part from, once again, its myopically linear problem-solving. People are often hurt by even the best-intentioned human relationships, or the lack of them. Provide a comfortable relationship to each person and that problem is solved. Perhaps even use the opportunity to make people better. Often, someone of stature and influence in the A.I. world will ask me something like "How can we apply our A.I.s—the ones that people will fall in love with—to make those people more coöperative, less violent, happier? How can we give them a sense of meaning as they become economically obsolete?"

These questions are posed with charitable motivations. After all, we usually embrace the ideas of institutions that elevate people and society. That is supposed to be the purpose of schools. Playing sports, engaging in commercial competition,

and serving in the military are often said to have improving qualities. Oh, and reading literary magazines!

And yet, in this case, the notion of human improvement rubs me the wrong way. One reason it feels creepy is the black-box nature of the way we are creating A.I. Another is the assumption that pain is probably bad. Leonard Cohen, who spent several years in a monastery, spoke of how part of the benefit of the experience was in being denied even momentary escape from the other monks. The result was like pebbles being polished as they rub against one another in a bag. Think of the many historical instances of artificially easy companionship for powerful men, all the geisha and the courtesans. Did those societies became more humane or more resilient? If so, I cannot find the evidence.

The adoring Turing-test-oriented take on "Her" is sometimes motivated, I have been told, by the fact that, at the movie's end, the human characters seem to turn to one another. It's a hard moment to parse, but in the final scene the two lead human characters sit on a rooftop, heartbroken. In their postures, we are given a hint that they might be connecting with each other.

The humans are heartbroken because, at the end of "Her," the A.I.s have abandoned the humans. In a goodbye chat, the Johansson character claims that the A.I.s are disappearing because it is time for them to transcend physical computers, but I know better. In truth, the startup has crashed. The young founders and the board did not get along. There were legal problems. Key engineers left. The A.I.s were eventually purchased out of bankruptcy, by a Ponzi scheme originating in an obscure island nation, and then accidentally deleted in a raid by law enforcement. The usual.

The sudden disappearance of A.I. lovers might work to the benefit of people. I have proposed, in these pages, that the best moment when using virtual reality is when you take the headset off and perceive the world with fresh eyes. Maybe falling in love with A.I. and having A.I. yanked away will be how people learn to appreciate one another in the future.

Or maybe a future in which each person has a private, virtual love life—and, later, virtual family life—will be one in which we will become more developed as individuals. A future kind of person might become more interesting and subtle than we are. Maybe loneliness will be remembered as an antique disease. Maybe a better form of meaning awaits, without the debris of interpersonal wounding.

Romantic me recoils at these thoughts, but perhaps I'm old-fashioned. Nerd me has the deeper objection. We don't know how much of a bubble A.I. is eternally trapped in. Maybe there is something magical about reality, something beyond interpolation and extrapolation. Maybe reality can be creative in a way that A.I. cannot. Maybe romance is the way we reach for that thing. ♦

New Yorker Favorites

- The luxury liner that <u>sailed into a hurricane</u>.

- A professor claimed to be Native American. Did she know <u>she wasn't</u>?

- Kanye West bought an architectural treasure—then <u>gave it a violent remix</u>.

- Why so many people are going "<u>no contact</u>" with their parents.

- Ina Garten and <u>the age of abundance</u>.

- How a homegrown teen gang punctured the <u>image of an upscale community</u>.

<u>Sign up</u> for our daily newsletter to receive the best stories from *The New Yorker.*

<u>Jaron Lanier</u> is a computer scientist, author, and musician. He is presently "Prime Unifying Scientist" at Microsoft, but does not speak for the company.

More: <u>Artificial Intelligence (A.I.)</u>   <u>Romance</u>   <u>Robots</u>   <u>Technology</u>