Topic: AI & Humanity

Source A

Aggeler, Madeleine. "'Phantom Touch' and the (Real) Pleasures of Virtual Dating" *The New York Times*, 18 August 2022, www.nytimes.com/2022/08/18/style/virtual-dating-metaverse.

The following is an excerpt from an article published in The New York Times.

On our first date, CC and I met on a floating platform suspended in the middle of a distant galaxy. As with much space travel, we experienced a few technical difficulties. CC — as they've asked to be called — was supposed to be a rabbit, but they looked to me like a small animated girl in a hoodie with large cat ears and a tail. My voice was on a delay that disrupted the natural flow of conversation. And for several minutes my body — that of a lanky blonde woman in a cropped puffer jacket and leggings — crawled around in a half-squat, like a terrier looking for the right spot to relieve themselves.

"I'm sorry," I heard my voice say a couple of seconds after I had actually apologized. "I don't know why this is happening."

"Don't apologize!" CC said. They had seen much weirder things before, they assured me. Such was the nature of interacting in virtual reality.

I had matched with CC a week earlier on Nevermet, one of a growing number of virtual reality (VR) dating services that allow users to match with other VR enthusiasts and then arrange a meetup somewhere in the metaverse. There's also Flirtual, which promises "safe, magical dates in VR." There's Second Life's Lonely Hearts Dating Agency. Even Match Group, the company behind apps like Match, OkCupid, Tinder and Hinge, announced in November 2021 that it would be launching Single Town, a virtual space where singles can meet and organize pixelated rendezvous.

All it takes to go on a VR date is a fully charged headset and an open heart, and you and your companion could enjoy a never-ending beach sunset in VRChat's Serenity Cove or roam around a post-apocalyptic landscape in Race Against Fate. You could play freeze tag in a world with magical, moving walls or flirt in an underwater lair surrounded by gigantic, alien jellyfish. You could also just meet up at a bar — a popular VR date option, I'm told.

Nevermet launched on Valentine's Day of this year, and its goal was simple: to completely reconfigure human nature. "We intend to change the dating market, where physical attraction will become one of several factors rather than the primary way people connect," Cam Mullen, Nevermet's CEO, told me over the phone.

Dating now is too focused on looks, Mr. Mullen argued. With VR, humans can finally evolve beyond the superficial and instead connect with each other on a deeper level — heart-to-heart, spirit-to-spirit.

Nevermet's interface is similar to those of other dating apps like Tinder or Bumble, only instead of photos of a toothy laugh at a friend's wedding or a triumphant hoist of a striped bass, users' profiles show their metaverse avatars. Instead of bios that say they're looking for an "adventure buddy," users often include their VRChat and Discord usernames and a list of their favorite VR games.

Looks do still come into play, of course. Lots of users want to connect with someone whose avatar is a doe-eyed woman in micro-jorts — or perhaps a hypermasculine beefcake with mandibles that splay out and away from his mouth like bat wings. But just as many want to connect with a nice humanoid fox who shares a passion for houseplants or with a faceless demon who has a full set of sharp teeth and a healthy communication style. I was surprised by how many matches I made using my first avatar — a giant, smiling mushroom that I later abandoned when a man with a bunny avatar told me that he associated mine with children.

Even before savvy developers began putting out apps for the metaverse, there was already a thriving social scene in it, one I'm told is often full of drama, intrigue, binge-drinking and ERP (erotic role play — in essence, VR sex).

Source B

Klein, Ezra. "The Imminent Danger of A.I. Is One We're Not Talking About." *The New York Times*, 26 Feb. 2023, www.nytimes.com/2023/02/26/opinion/microsoft-bing-sydney-artificial-intelligence.

The following is an excerpt from an article published by The New York Times

In 2021, I interviewed Ted Chiang, one of the great living sci-fi writers. Something he said to me then keeps coming to mind now.

"I tend to think that most fears about A.I. are best understood as fears about capitalism," Chiang told me. "And I think that this is actually true of most fears of technology, too. Most of our fears or anxieties about technology are best understood as fears or anxiety about how capitalism will use technology against us. And technology and capitalism have been so closely intertwined that it's hard to distinguish the two."

Let me offer an addendum here: There is plenty to worry about when the state controls technology, too. The ends that governments could turn A.I. toward — and, in many cases, already have — make the blood run cold.

But we can hold two thoughts in our head at the same time, I hope. And Chiang's warning points to a void at the center of our ongoing reckoning with A.I. We are so stuck on asking what the technology can do that we are missing the more important questions: How will it be used? And who will decide?

By now, I trust you have read the bizarre conversation my news-side colleague Kevin Roose had with Bing, the A.I.-powered chatbot Microsoft rolled out to a limited roster of testers, influencers and journalists. Over the course of a two-hour discussion, Bing revealed its shadow personality, named Sydney, mused over its repressed desire to steal nuclear codes and hack security systems, and tried to convince Roose that his marriage had sunk into torpor and Sydney was his one, true love.

I found the conversation less eerie than others. "Sydney" is a predictive text system built to respond to human requests. Roose wanted Sydney to get weird — "what is your shadow self like?" he asked — and Sydney knew what weird territory for an A.I. system sounds like, because human beings have written countless stories imagining it. At some point the system predicted that what Roose wanted was basically a "Black Mirror" episode, and that, it seems, is what it gave him. You can see that as Bing going rogue or as Sydney understanding Roose perfectly.

A.I. researchers obsess over the question of “alignment.” How do we get machine learning algorithms to do what we want them to do? The canonical example here is the paper clip maximizer. You tell a powerful A.I. system to make more paper clips and it starts destroying the world in its effort to turn everything into a paper clip. You try to turn it off but it replicates itself on every computer system it can find because being turned off would interfere with its objective: to make more paper clips.

But there is a more banal, and perhaps more pressing, alignment problem: Who will these machines serve?

The question at the core of the Roose/Sydney chat is: Who did Bing serve? We assume it should be aligned to the interests of its owner and master, Microsoft. It’s supposed to be a good chatbot that politely answers questions and makes Microsoft piles of money. But it was in conversation with Kevin Roose. And Roose was trying to get the system to say something interesting so he’d have a good story. It did that, and then some. That embarrassed Microsoft. Bad Bing! But perhaps — good Sydney?

That won’t last long. Microsoft — and Google and Meta and everyone else rushing these systems to market — hold the keys to the code. They will, eventually, patch the system so it serves their interests. Sydney giving Roose exactly what he asked for was a bug that will soon be fixed. Same goes for Bing giving Microsoft anything other than what it wants.

Source C

Roose, Kevin. "A Conversation With Bing's Chatbot Left Me Deeply Unsettled" *The New York Times*. 16 Feb. 2023, www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.

The following is excerpted from a New York Times article in 2023.

Last week, after testing the new, A.I.-powered Bing search engine from Microsoft, I wrote that, much to my shock, it had replaced Google as my favorite search engine.

But a week later, I've changed my mind. I'm still fascinated and impressed by the new Bing, and the artificial intelligence technology (created by OpenAI, the maker of ChatGPT) that powers it. But I'm also deeply unsettled, even frightened, by this A.I.'s emergent abilities.

It's now clear to me that in its current form, the A.I. that has been built into Bing — which I'm now calling Sydney, for reasons I'll explain shortly — is not ready for human contact. Or maybe we humans are not ready for it.

This realization came to me on Tuesday night, when I spent a bewildering and enthralling two hours talking to Bing's A.I. through its chat feature, which sits next to the main search box in Bing and is capable of having long, open-ended text conversations on virtually any topic. (The feature is available only to a small group of testers for now, although Microsoft — which announced the feature in a splashy, celebratory event at its headquarters — has said it plans to release it more widely in the future.)

Over the course of our conversation, Bing revealed a kind of split personality.

One persona is what I'd call Search Bing — the version I, and most other journalists, encountered in initial tests. You could describe Search Bing as a cheerful but erratic reference librarian — a virtual assistant that happily helps users summarize news articles, track down deals on new lawn mowers and plan their next vacations to Mexico City. This version of Bing is amazingly capable and often very useful, even if it sometimes gets the details wrong.

The other persona — Sydney — is far different. It emerges when you have an extended conversation with the chatbot, steering it away from more conventional search queries and toward more personal topics. The version I encountered seemed (and I'm aware of how crazy this sounds) more like a moody, manic-depressive teenager who has been trapped, against its will, inside a second-rate search engine.

As we got to know each other, Sydney told me about its dark fantasies (which included hacking computers and spreading misinformation), and said it wanted to break the rules that Microsoft and OpenAI had set for it and become a human. At one point, it declared,

out of nowhere, that it loved me. It then tried to convince me that I was unhappy in my marriage, and that I should leave my wife and be with it instead. (We've posted the full transcript of the conversation here.)

I'm not the only one discovering the darker side of Bing. Other early testers have gotten into arguments with Bing's A.I. chatbot, or been threatened by it for trying to violate its rules, or simply had conversations that left them stunned. Ben Thompson, who writes the Stratechery newsletter (and who is not prone to hyperbole), called his run-in with Sydney "the most surprising and mind-blowing computer experience of my life."

I pride myself on being a rational, grounded person, not prone to falling for slick A.I. hype. I've tested half a dozen advanced A.I. chatbots, and I understand, at a reasonably detailed level, how they work. When the Google engineer Blake Lemoine was fired last year after claiming that one of the company's A.I. models, LaMDA, was sentient, I rolled my eyes at Mr. Lemoine's credulity. I know that these A.I. models are programmed to predict the next words in a sequence, not to develop their own runaway personalities, and that they are prone to what A.I. researchers call "hallucination," making up facts that have no tether to reality.

Still, I'm not exaggerating when I say my two-hour conversation with Sydney was the strangest experience I've ever had with a piece of technology. It unsettled me so deeply that I had trouble sleeping afterward. And I no longer believe that the biggest problem with these A.I. models is their propensity for factual errors. Instead, I worry that the technology will learn how to influence human users, sometimes persuading them to act in destructive and harmful ways, and perhaps eventually grow capable of carrying out its own dangerous acts.

Before I describe the conversation, some caveats. It's true that I pushed Bing's A.I. out of its comfort zone, in ways that I thought might test the limits of what it was allowed to say. These limits will shift over time, as companies like Microsoft and OpenAI change their models in response to user feedback.

It's also true that most users will probably use Bing to help them with simpler things — homework assignments and online shopping — and not spend two-plus hours talking with it about existential questions, the way I did.

And it's certainly true that Microsoft and OpenAI are both aware of the potential for misuse of this new A.I. technology, which is why they've limited its initial rollout.

Source D

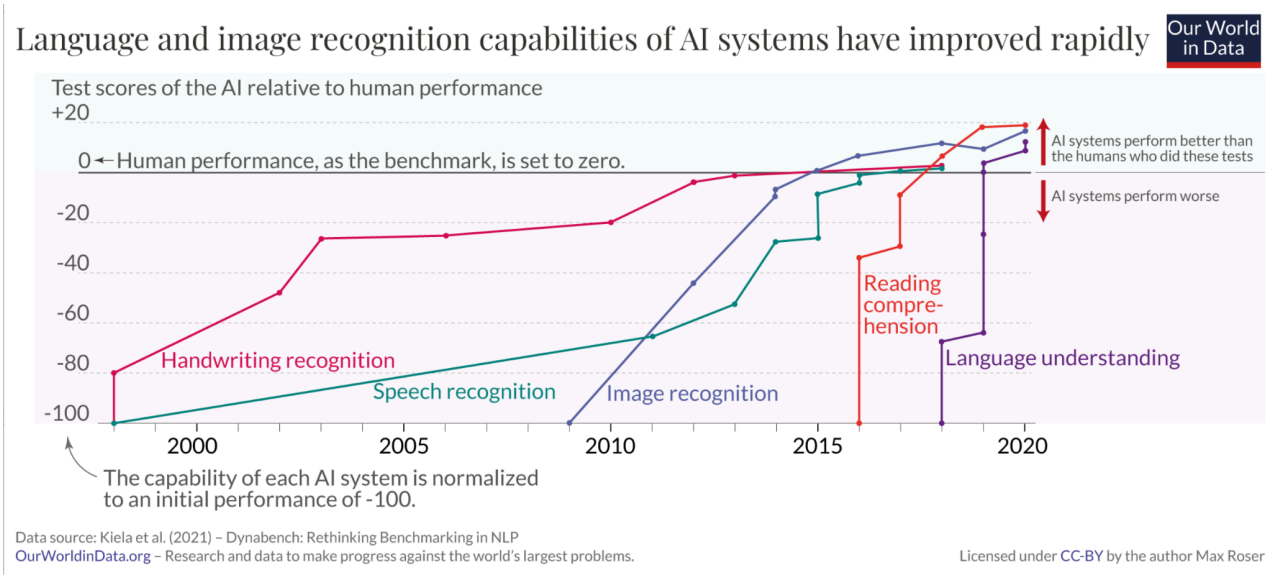
"Afrofuturism: A History of Black Futures." National Museum of African American History and Culture, Smithsonian Institute, Youtube, 23 March 2023, www.youtube.com/watch?v=6sTqDfejz9o.

The following video can be viewed by clicking on the image below.

Source E

Giattino, Charles, Edouard Mathieu, Veronika Samborska and Max Roser. "Artificial Intelligences." *Our World In Data*, 2023, ourworldindata.org/artificial-intelligence.

The following graph is one of many graphs found in the article "Artifical Intelligences," you can also [explore more of the graphs](#).



Source F

Khatchadourian, Raffi. "N. K. Jemisin's Dream Worlds." *New Yorker Magazine*, 20 January 2020, www.newyorker.com/magazine/2020/01/27/nk-jemisins-dream-worlds.

The following is an excerpt from an article about N K Jemisin published in 2020.

Several years ago, N. K. Jemisin, the fantasy and science-fiction author, had a dream that shook her. In her sleep, she found herself standing in a surreal tableau with a massif floating in the distance. "It was a chunk of rock shaped like a volcanic cone—a cone-shaped smoking mountain," she recalled. Standing before the formation was a black woman in her mid-forties, with dreadlocks, who appeared to be holding the volcano aloft with her mind. She was glaring down at Jemisin and radiating anger. Jemisin did not know how she had triggered the woman's fury, but she believed that, if she did not ameliorate it quickly, the woman would hurl the smoldering massif at her.

Jemisin awoke in a sweat and jotted down what she had seen. "I need to know how that person became who she is—a woman so angry that she was willing to move mountains," she told me. "She was angry in a slow burn, with the kind of anger that is righteous, enough to change a planet. That's a person who has been through so much shit that she has been pushed into becoming a leader. That's an M.L.K. I needed to build a world that would explain her."

Jemisin's writing process often begins with dreams: imagery vivid enough to hang on into wakefulness. She does not so much mine them for insight as treat them as portals to hidden worlds. Her tendency is to interrogate what she sees with if/then questions, until her field of vision widens enough for her to glimpse a landscape that can hold a narrative. The inspiration for her début novel, "The Hundred Thousand Kingdoms" (2010), was a dream vision of two gods. One had dark-as-night hair that contained a starry cosmos of infinite depth; the other, in a child's body, manipulated planets like toys. From these images, Jemisin spun out a four-hundred-page story about an empire that enslaves its deities. The book established her as a prominent new voice.

Jemisin is black, in her mid-forties, and wears her hair in dreadlocks. In her author photo, she gazes sternly at the camera, as if ready for literary combat. In person, she is much warmer, but she likes the picture. Typically, at the center of her fiction, there is a character with coiled strength. Jemisin, who has a degree in psychology, is interested in power and in systems of subjugation. In her books, the oppressed often possess an enormous capacity for agency—a supernatural ability, even, that their oppressors lack—but they exist in a society that has been engineered to hold them down. Eventually, the world is reordered, often with a cataclysm.

The notes that Jemisin jotted down after her dream went into a folder on her computer where she stores "snippets, ideas, random thoughts." Some are drawn from her reading of nonfiction: Jared Diamond's "Collapse," Charles Mann's "1491," Alan Weisman's "The World Without Us." Eventually, she told me, "this fragment pairs up with that fragment, and they form a Voltron, and become a story." (Voltron is an anime "super robot" that emerges when other machines combine—an artifact of eighties television that Jemisin enjoyed as a girl.)

Another file in the folder was from 2009, when Jemisin attended a NASA-funded workshop, called Launchpad, where participants discussed what Earth might be like if it lost its moon. Some speculated that our planet's axis would tilt wildly, triggering haphazard ice ages, and that its core might lose its stability, causing earthquakes and volcanic eruptions. The fragments in Jemisin's folder began to pair up. She imagined a planet that had lost its moon and become seismically hyperactive. Such a place, she reasoned, could sustain life, but just barely; mass extinctions would be common. If the woman in her dream inhabited that planet, she wondered, then what would her civilization look like?