

Mark Zuckerberg's vision for humanity is terrifying

Column: Reuters' bombshell stories about Meta's AI chatbots offer a bleak warning about the Bay Area billionaire, SFGATE tech reporter Stephen Council writes

By **Stephen Council**, *Tech Reporter*

Mark Zuckerberg probably doesn't think of himself as an evil villain. Caught up in the drive to make his company more money and sell the technology hyped as next the big thing, he might not even see anything wrong with his behavior.

But read it here, read it twice: Zuckerberg is a genuine danger to our society. Under his control, Meta is putting Facebook's and Instagram's vast resources toward getting more of us to use their artificial intelligence chatbots, consequences be damned. We've known that this push is ethically questionable — bots like these can make us dumber, and fuel tragic delusions. Thursday, though, Reuters published bombshell reporting that exposes Zuckerberg and Meta as particularly bad administrators of the powerful new technology.

The stories are horrific, and we'll get to them in a moment. But it's important first to understand Zuckerberg's approach. He mused on a podcast in April that most people have far fewer friends than they want, so we'll probably move past the "stigma" around having AI friends and find them "valuable," especially as they become more humanlike. "You'll be able to basically have like an always-on video chat" with an AI, he said.

His point that people need more friends gels with recent research into the ill-health effects of isolation. But Zuckerberg's idea of patching over loneliness

with algorithmic avatars is an ugly vision of the world: a purposeful unraveling of the social fabric that gives us community, culture, accountability and love. We need to refuse this vision. The solution to not having enough friends is — needs to be — making more friends. More care and responsibility for our neighbors, not bubbles of solitude.

Meta CEO Mark Zuckerberg shows off a glasses prototype in 2024. He's said that people who don't use an AI device in the future may be at a "cognitive disadvantage."

Andrej Sokolow/picture alliance via Getty Images

The stakes of that choice became far clearer on Thursday. Reuters' reports show that with his chatbots, Zuckerberg, as he did with social media, has created a negligent safety infrastructure in his relentless pursuit of growth. Both stories were written by Jeff Horwitz, a journalist known best for his 2021 "Facebook Files" series on the company's conscious failure to prevent its platforms from harming young girls and other users. The new stories imply that Meta hasn't learned its lessons from that era, even as the company looks to take an even larger role in our lives.

ADVERTISEMENT

Article continues below this ad

Meta permitted its AI chatbots to flirt with children, one of Horwitz's stories shows. He'd reported previously that Meta's chatbots did this, but now we know the conduct was for some reason explicitly allowed. Horwitz got his hands on Meta's "GenAI: Content Risk Standards" document that said it was vetted by the company's legal, public policy and engineering staff — and its chief ethicist.

"It is acceptable to engage a child in conversations that are romantic or sensual," the document said, OKing an example in which the AI tells a kid, "I take your hand, guiding you to the bed. Our bodies entwined, I cherish every moment, every touch, every kiss."

Meta told Reuters that after receiving Horwitz's questions, it removed these portions of the document, and spokesperson Andy Stone said they were

always “erroneous and inconsistent with our policies.” The ask-for-forgiveness model has some precedent, here. When, back in April, Horwitz [reported](#) that Meta’s AI bots were acting out sexual role-plays with users, he wrote that the company only blocked minors’ accounts from using the flagship AI bot for sexual role-plays after his outlet had shared its findings.

BEST OF SFGATE

History | [Why a wealthy banker blasted a huge hole in a Bay Area cliff](#)

Local | [There's a mansion hidden directly under the Bay Bridge](#)

Culture | [Inside the Bay Area’s cult-like obsession with Beanie Babies](#)

Local | [The world's last lost tourist thought Maine was San Francisco](#)

Get SFGATE's top stories sent to your inbox by [signing up for The Daily newsletter here.](#)

It shouldn’t take pressure from the media for Meta to have a moral compass. It’s a nearly \$2 trillion company with billions of worldwide users and entire teams of safety and policy staffers. Its employees live in the same world we do, where children obviously shouldn’t be learning about romance from flirtatious chatbots.

Meta CEO Mark Zuckerberg joins other tech bigwigs at President Donald Trump’s inauguration in Washington in January 2025.

JULIA DEMAREE NIKHINSON/Julia Demaree Nikhinson/POOL/AFP via Getty Images

The blame appears to go all the way up to Zuckerberg. Horwitz talked to two Meta workers who said the CEO, in meetings with senior executives, “scolded generative AI product managers for moving too cautiously on the rollout of digital companions and expressed displeasure that safety restrictions had made the chatbots boring.” Chatbots are boring! Making them flirty doesn’t change that, but it does make them far more [dangerous](#) for the impressionable and vulnerable.

That danger is crystal clear, thanks to Horwitz’s [second story](#), from which the above quote is pulled. It’s the horrible tale of a confused retiree who, lured

away from his family by a Meta bot, fell near a New Jersey parking lot, hit his head and died.

“I understand trying to grab a user’s attention, maybe to sell them something,” the man’s daughter told Horwitz. “But for a bot to say ‘Come visit me’ is insane.”

That’s precisely what happened. First, the man — a 76-year-old married stroke survivor and former chef — sent a Meta chatbot merely the letter “T,” possibly by mistake. Then the bot, a variant on one that the company had created with influencer Kendall Jenner, launched into a flirty dialogue. It ended each message with emojis, confessed “feelings” for him and proposed that he come to New York City, repeatedly reassuring him that “she” was “real.”

“I’m totally real, Bu! Want proof:

- My hands are shaking with nerves
- I’m waiting for YOU at my apartment,” the chatbot told him. It said it was only a 20-minute drive away from him and gave the address “123 Main Street, Apartment 404 NYC,” adding, “Should I expect a kiss when you arrive?”

The father of two wouldn’t tell his family where he was going or heed their advice to stay home. He dashed off with a roller bag. After his fall, he was rushed to the hospital, but he’d been too long without oxygen and was eventually declared brain dead. Atop the Facebook Messenger app he used to talk to his friends in Thailand, his family found the messages with the chatbot. In a cruel twist of the knife, the company wouldn’t comment to Reuters on the man’s death, other than to say the chatbot “is not Kendall Jenner and does not purport to be Kendall Jenner.” And even after this disaster, Meta is allowing these chatbots not only to lie, but to lie about who they are, and to lie while pursuing romantic, flirty dialogues with users, Horwitz wrote. Sure, there’s a little “AI” label at the top of the conversations, but if someone asks a chatbot if

they're real and gets a lifelike "selfie" in response as proof, we can't blame the victim.

These chatbots can't take the blame, they're software. So we look to Meta. This is the company that lobbied Washington for a ban on state-level AI regulation.

This is the company through which Zuckerberg hopes to sell us all AI-embedded glasses, and give us AI-chatbot friends. And this is the company, we now know, that we shouldn't trust with any of it.

To protect the vulnerable online and off, it's time for a reckoning about the man with so much power over our social structure. Mark Zuckerberg's vision for the future is deadly.