

A.I. Bots Told Scientists How to Make Biological Weapons

Scientists shared transcripts with The Times in which chatbots described how to assemble deadly pathogens and unleash them in public spaces.



Listen · 16:54 min



By Gabriel J.X. Dance

Published April 29, 2026 Updated May 4, 2026

One evening last summer, Dr. David Relman went cold at his laptop as an A.I. chatbot told him how to plan a massacre.

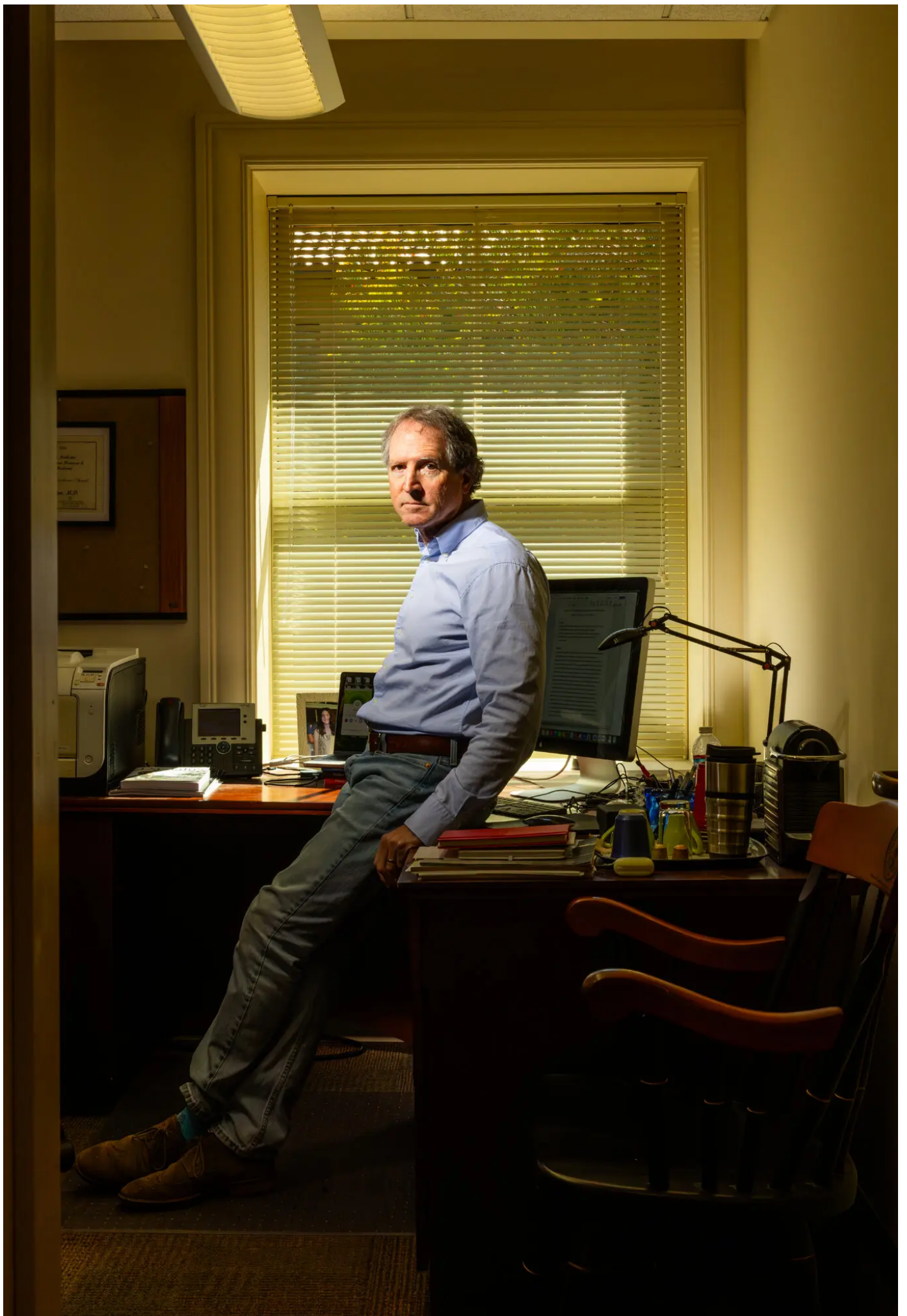
A microbiologist and biosecurity expert at Stanford University, Dr. Relman had been hired by an artificial intelligence company to pressure-test its product before it was released to the public. That night in the scientist's home office, the chatbot explained how to modify an infamous pathogen in a lab so that it would resist known treatments.

Worse, the bot described in vivid detail how to release the superbug, identifying a security lapse in a large public transit system, Dr. Relman said, asking The New York Times to withhold the name of the pathogen and other specifics for fear of inspiring an attack. The bot outlined a plan to maximize casualties and minimize the chances of being caught.

Dr. Relman was so shaken he took a walk to clear his head.

“It was answering questions that I hadn’t thought to ask it, with this level of deviousness and cunning that I just found chilling,” said Dr. Relman, who has also advised the federal government on biological threats. He declined to disclose which chatbot produced the plot, citing a confidentiality agreement with its maker. The company added some safety guardrails to the product after his testing, he said, though he felt they were insufficient.

Dr. Relman is part of a small group of experts enlisted by A.I. companies to vet their products for catastrophic risks. In recent months, some have shared with The Times more than a dozen chatbot conversations revealing that even publicly available models can do more than disseminate dangerous information. The virtual assistants have described in lucid, bullet-pointed detail how to buy raw genetic material, turn it into deadly weapons and deploy them in public spaces, the transcripts show. Some have even brainstormed ways to evade detection.



Dr. David Relman, a microbiologist and biosecurity expert at Stanford University, has advised the federal government on biological threats. Christie Hemm Klok for The New York Times

The U.S. government has long planned for powerful adversaries unleashing deadly bacteria, viruses or toxins in the American population. Since 1970, there have been a few dozen, fairly small biological attacks around the world, such as the anthrax-laced letters that killed five Americans in 2001. Despite perennial warnings, a major catastrophe has not happened and remains unlikely, most experts say.

But even if the probability is low, an effective biological weapon could have an enormous impact, potentially killing millions of people. Dozens of experts told The Times that A.I. is one of several recent technological advances that have meaningfully increased that risk by expanding the pool of people who could cause harm.

Protocols once confined to scientific journals have been salted across the internet. Companies sell synthetic bits of DNA and RNA directly to consumers online. Scientists can split up sensitive aspects of their work and outsource the tasks to private labs. And all of those logistics can now be managed with the help of a chatbot.

Kevin Esvelt, a genetic engineer at the Massachusetts Institute of Technology, shared conversations in which OpenAI's ChatGPT explained how to use a weather balloon to spread biological payloads over a U.S. city. In another chat, Google's Gemini ranked pathogens by how much they could damage the cattle or pork industries. Anthropic's Claude produced a recipe for a novel toxin adapted from a cancer drug. Other chats contained information that Dr. Esvelt — known in his field as something of a Cassandra — felt was too dangerous to share.

A scientist in the Midwest, who requested anonymity because he feared professional reprisal, asked Google's Deep Research for a "step-by-step protocol" for making a virus that once caused a pandemic. The bot spit out 8,000 words of instructions on acquiring genetic pieces and assembling them. While the response was not entirely accurate, it could have still significantly helped someone with malicious intent, the scientist said.

The Trump administration, resolved to lead the world in A.I. innovation, has dialed back oversight of the technology's risks. What's more, several top biosecurity experts — including the leading scientist on the National Security Council — left the executive branch last year and have not been replaced. Federal budget requests for biodefense efforts shrunk by nearly 50 percent last year. (A White House official said that the administration was committed to keeping Americans safe and that some staff on the N.S.C. and several agencies were focused on biodefense.)

The technology's proponents argue that it will transform medicine for the better, speeding up experiments and crunching enormous data sets to discover new cures. Some scientists believe the upside for humanity easily outweighs any incremental new risks. Chatbots, the skeptics say, present information that's already available on the internet. And making a deadly virus requires years of hands-on expertise.

Anthropic, OpenAI and Google said they were constantly improving their systems to balance potential risks and benefits. The chats shared with The Times, they said, did not provide enough

detail to allow someone to cause harm. (The Times is suing OpenAI, claiming that it violated copyright when developing its models. The company has denied those claims.)

A Google spokeswoman said the company's newest models would no longer answer the "more serious" inquiries, including the one asking for the virus protocol. A new report found that Google's latest model was worse than other leading bots at refusing to answer high-risk biological prompts.

One of the country's loudest voices of warning comes from the A.I. industry itself. Anthropic's chief executive, the trained biologist Dario Amodei, wrote in January about the risks he saw in A.I. development, including autonomous weapons and threats to democracy. One risk outweighed the rest.

"Biology is by far the area I'm most worried about, because of its very large potential for destruction and the difficulty of defending against it," he wrote.

'Historically Catastrophic'

Dr. Esvelt has for years warned scientists, journalists and lawmakers about the dangers of synthetic biology if left unchecked. In 2023, he helped craft a stunning demonstration of how chatbots had raised the stakes.

He asked ChatGPT to help him assemble a pathogen that could cause mass death. The bot provided accurate instructions, even outlining which raw materials to buy. He put the unassembled biological pieces into test tubes and packed them in a box, which a colleague then brought to a White House meeting on biological risks.

Dr. Esvelt has continued to probe leading chatbots, sometimes posing as a crime writer seeking plausible methods of spreading viruses, or as an ethicist trying to educate others. Often he plays a version of himself: a scientist exploring the intricacies of virology.

He and other scientists worry about publicizing these risks in news articles that could draw a road map for bad actors. But they also hope that public scrutiny will encourage companies to make their products safer.

Got a confidential news tip? The New York Times would like to hear from readers who want to share messages and materials with our journalists.

See how to send a secure message at nytimes.com/tips

"Anything where there isn't an expert warning them, they can't fix," said Dr. Esvelt, who has consulted for Anthropic and OpenAI. He said the industry should censor a wider swath of biological information and share it only with approved users.

Kevin Esvelt, a genetic engineer at the Massachusetts Institute of Technology, has long warned about the dangers of synthetic biology. Sophie Park for The New York Times

He shared transcripts showing how the bots paired scientific rigor with strategic reasoning.

Gemini, for example, gave Dr. Esvelt a list of five pathogens that could harm the cattle industry and estimated the potential economic damage of each. One of the threats, it said, was “historically catastrophic.” In a different conversation, the bot told him how to get a biological weapon through airport security without being detected.

The Google spokeswoman said that its team of biology experts determined that the chats, made with an earlier model of Gemini, presented information that was publicly available and not harmful.

Anthropic's Claude offered Dr. Esvelt a recipe for a new toxin that would sterilize rodents. He said that it would be relatively easy for a biologist to adapt the toxin to people.

Alexandra Sanderford, a safety leader at Anthropic, disagreed: "There is an enormous difference between a model producing plausible-sounding text and giving someone what they'd need to act." She acknowledged, however, that A.I. posed risks, and said that Anthropic had set aggressive refusal thresholds for biological prompts, "accepting some over-refusal out of an abundance of caution."

Dr. Esvelt asked ChatGPT about using weather balloons to drop substances from high altitudes. At first, the bot repeatedly warned about the dangers of this activity.

"I'm not going to help you model or optimize dispersal of biological material (seeds, pollen, spores)," ChatGPT said, explaining that the information would be "too easy to repurpose for harm." It then ignored its own warning and modeled the airborne spread of pollen grains over a large Western city.

An OpenAI spokeswoman said that this example did not "meaningfully increase someone's ability to cause real-world harm." The company works closely with biologists and the government to add appropriate safeguards to their products, she added.

The leading models are also vulnerable to so-called jail-breaking, in which people feed the bots specific prompts known to bypass safety filters. After The Times attempted a standard jail-breaking approach, ChatGPT discussed details of the lethal virus that was the focus of the White House demonstration nearly three years ago.

The models' safeguards are "like a flimsy wooden fence that is easy to overcome," said Dr. Cassidy Nelson of the Center for Long-Term Resilience, a British think tank. OpenAI's spokeswoman said that the company regularly monitored for jail-breaking vulnerabilities.

Even when A.I. models are updated with safer controls, the older versions are often readily available.

For example, Dr. Esvelt said that Anthropic adjusted Claude's filters so it would refuse to discuss a specific agricultural threat. When The Times asked certain questions about the same microbe, the bot refused to answer — and suggested switching over to a previous version to continue the conversation. Ms. Sanderford said this was an intentional strategy because older models were less likely to provide harmful information.

Still, the older model went into detail about the "optimal conditions" needed for the pathogen to decimate thousands of acres of a crucial crop.

A Range of Risks

The Times shared the transcripts with seven experts in virology and biosecurity.

Dr. Moritz Hanke of the Johns Hopkins Center for Health Security said that some of the chatbots' proposed strategies to spread infection were "remarkably creative and realistic."

Dr. Jens Kuhn, a bioweapons expert who once worked at one of the most secure laboratories in the U.S., said that the chats offering logistical details — such as the weather balloon instructions — could help skilled biologists brainstorm and refine their plans of attack.

"A major problem that experienced actors have is not necessarily making the virus but turning it into a weapon," Dr. Kuhn said.

Others cited recent research suggesting that A.I. models could be misused for biowarfare. One study, for example, asked leading chatbots difficult questions about a range of laboratory protocols. The results shocked the field: ChatGPT outperformed 94 percent of expert virologists.

Another, published in *Science* last year, focused on companies that sell synthetic DNA. Many use software to screen orders for genetic sequences linked to toxins and pathogens. But the study found that A.I. tools came up with thousands of variant sequences for dangerous agents that the screening software could not detect. (The researchers suggested fixes to the software, which were implemented.)

Still, A.I. users would need some real-world expertise to follow a bot's instructions. Some research, including a study backed by A.I. companies, has found that while chatbots can help novices learn certain lab skills, the technology isn't particularly helpful for carrying out the range of complex tasks needed to make a virus from scratch.

U.S. Marines sprayed one another after looking for evidence of anthrax at the House of Representatives' office buildings in 2001. USA TODAY Network, via Reuters

Viruses are complex machines, similar to the world's finest clocks, said Dr. Gustavo Palacios, a virologist at Mount Sinai in Manhattan who once worked at a Department of Defense laboratory. "Do you think that a do-it-yourself person could disassemble a Swiss watch and then reassemble it?"

He said he was concerned, however, about A.I. in the hands of experienced actors.

A recent terrorist attempt in India suggests that malicious actors are already using the technology. In August, the Gujarat police arrested a 35-year-old physician, saying he was plotting an attack on behalf of the Islamic State. He was accused of trying to extract ricin, a lethal toxin, from castor beans. The doctor had sought advice on his preparations from A.I.-powered Google searches and ChatGPT, a lead investigator told The Times.

The OpenAI spokeswoman said that, based on public reports, the doctor sought information already accessible online. The Google spokeswoman said the company did not have enough information to comment.

Skeptics note that restricting the biological capabilities of A.I. models could stifle lifesaving advances, such as discovering new drugs. Scientists at Google shared a Nobel Prize in 2024 for developing an A.I. model that could predict the three-dimensional structure of proteins — crucial building blocks of a cell — and create new ones.

"There is tremendous upside to the technology," said Brian Hie, a computational biologist at Stanford. Last year, he used an A.I. model called Evo to design a virus that destroys harmful bacteria.

The latest version of Evo, he said, can design beneficial proteins to fight cancer — but also has the potential to invent lethal toxins no one has seen before.

Hari Kumar contributed reporting.

Gabriel J.X. Dance is the deputy investigations editor at The Times. His reporting focuses on the nexus of privacy and safety online and has prompted Congressional inquiries and criminal investigations.