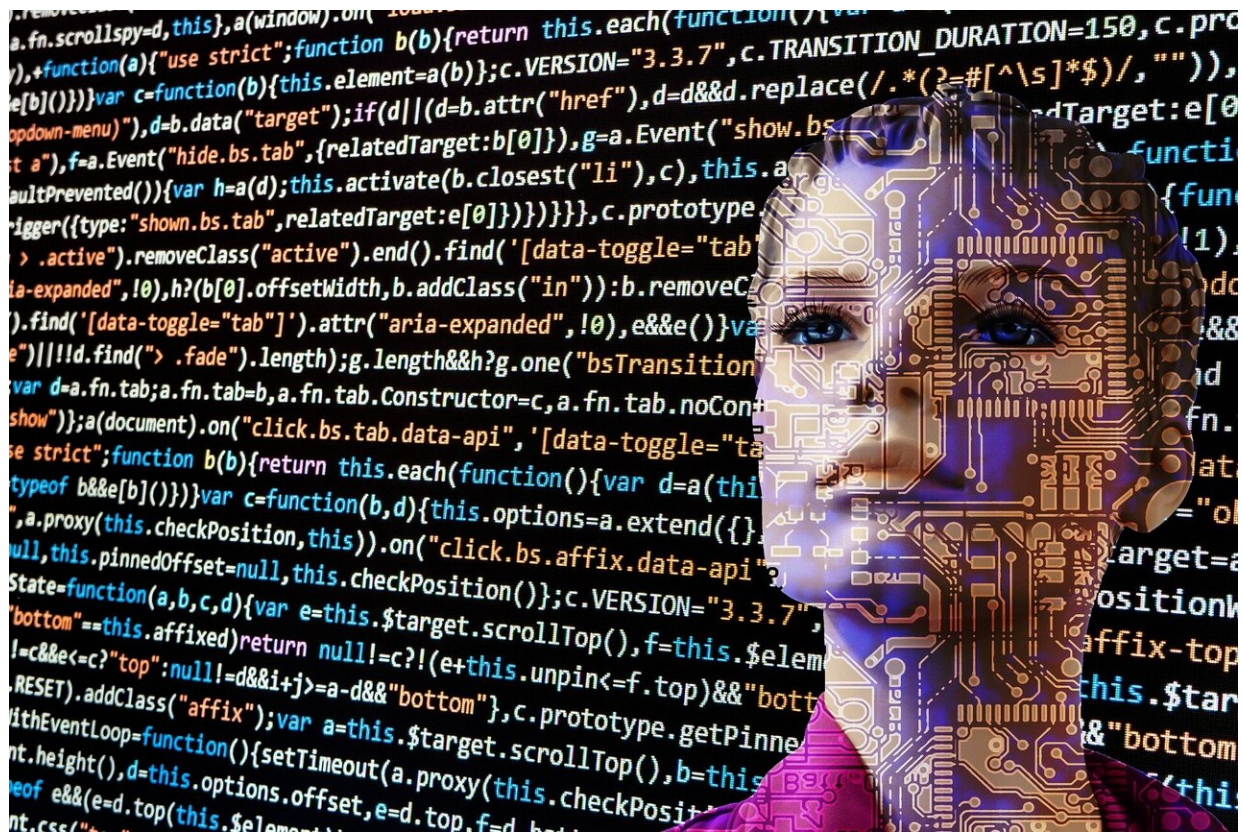


Q&A: Assessing the risks of existential terrorism and AI

September 28 2023, by J.T. Stone



Credit: CC0 Public Domain

Gary Ackerman, an associate professor and associate dean at the State University of New York's University at Albany College of Emergency Preparedness, Homeland Security and Cybersecurity (CEHC), has spent

decades studying terrorism around the world—from the motivations and capabilities of terrorist groups to the mitigation strategies governments use to defend against them.

Last month, Ackerman published [an article](#) in the *European Journal of Risk Regulation* that gained a substantial amount of media attention: "Existential Terrorism: Can Terrorists Destroy Humanity?" The paper, which Ackerman co-authored with Zachary Kallenborn of the Center for Strategic and International Studies (CSIS), explores the plausibility of terrorist organizations using emerging technologies such as AI to enact existential harm, including human extinction.

Ackerman has headed more than 10 large government-sponsored research projects over the past five years to address counterterrorism policy and operations, and has testified before the Senate Committee on Homeland Security about [terrorist](#) motivations for using [nuclear weapons](#). He is also a senior investigator and co-founder of the nation's first Center for Advanced Red Teaming (CART) housed at CEHC.

In this Q&A, Ackerman discusses existential terrorism and the threats it poses, what's being done to prevent the use of AI as a weapon, and why he found it necessary to publish an article about this topic now.

How do you define existential terrorism?

We define existential terrorism as terrorism that will cause sufficient harm to the continuation of humanity, either by wiping out the population completely or reducing it to an unviable quantity. Another understanding of existential risk that we discuss is the prevention of human flourishing, in which the human species gets stuck in a cycle where it cannot grow, such as in a global totalitarian society that oppresses all of mankind. But for the purposes of our research, we define existential terrorism as terrorism that brings about (or comes

close to bringing about) human extinction.

When people think about what could destroy humanity, they think of climate change, nuclear war or a pandemic, and not usually terrorism. Some people argue that terrorism at that scale is something seen only in science fiction or James Bond movies. We initially had the same reaction, but then we realized that no one has really taken this topic seriously. So we decided to take a more in-depth look at whether terrorists could ever cause a degree of harm that could put the existence of humanity in jeopardy.

How does emerging tech like AI contribute to the threat of existential terrorism?

It's really impossible for an individual or small group of terrorists to destroy humanity in most cases unless they have an extreme amount of leverage. One of the ways they can get leverage is through an enabling technology like AI, because it can act as a force multiplier, potentially even to cause harm at an extinction level. One example would be if terrorists hacked into an existing AI, say, that controlled nuclear weapons systems and set off nuclear war.

Another option would be if terrorists created a malevolent AI and instructed it to destroy humanity, although this option might be extremely difficult to do and remains highly speculative. This is because we don't yet have the kind of AI that could destroy humanity on its own, and we don't really know how far we are from that point—it could be five years, 50 years or maybe never.

The only current technology that terrorists could feasibly produce and deploy on their own to cause an [existential threat](#) is biotechnology. An example of this is if terrorists created a pandemic disease that was self-

replicating, extremely contagious and caused [high mortality rates](#), but this would require extremely high technical knowledge and specialized equipment. This explains why terrorists directly causing the end of humanity is very unlikely.

On the other hand, terrorists could cause harm indirectly by removing safeguards or preventing us from minimizing other risks. For example, terrorists could sabotage a rocket that we might send into space to divert a comet away from the Earth or remove safeguards that prevented an existing AI from going rogue. We call acts like these "spoilers," which we believe are much more plausible than terrorists causing existential harm directly. Fortunately, these require an existential risk to have already manifested on its own, which means that terrorists could not bring about this kind of harm completely on their own.

Why did you feel it was necessary to publish an article about this topic?

A lot of people dismiss these hypothetical scenarios as crazy or too far-fetched. Even if we find that there's not much of a threat, which is essentially what we have found to be the case at this moment, it's still worth considering such scenarios, so that we're prepared for future emerging threats, like AI. Even from this initial research, we now understand some of these emerging threats better and that there are some areas where existential harm from terrorists is feasible, such as in the case of spoilers.

The other reason we explored existential terrorism is that by exploring the most extreme scenarios, we can better calibrate the likelihood of less extreme cases of terrorism. Overall, we found that although there are definitely people who would like to destroy humanity, it's not something that I would lose sleep over at the moment. But, at some point, they

theoretically could succeed, so it's important to know what the threat might look like and what we can do to prevent it.

What's being done to prevent the potential use of AI as a weapon?

Not much has been done specifically to prevent AI from being used as a weapon on a human extinction scale. However, there's been a lot of work about AI risk and risk prevention published by think tanks like the Global Catastrophic Risk Institute (GCRI), where I'm also a senior advisor. In March, over 1,000 industry leaders, researchers and tech CEOs signed an [open letter](#) calling for a six-month moratorium on the development of advanced AI systems, citing AI's profound risks to society and humanity.

But most of the action taken by Congress, at least in the United States, has focused more on addressing the other risks of AI, like displacing jobs or being used by our adversaries to design better weapons. Very few people in our government are seriously looking at AI as an existential problem, even though people are slowly becoming aware of these potential threats. There's a legitimate worry that the smarter we make systems, even if they don't quite get to sentience, the more likely they could become a major risk.

Broadly speaking, we have to think of AI as a global issue. We may have disagreements with other countries, but neither Russia, China nor any of the United States' other rivals have any interest in the world being destroyed. When it comes to threats of existential terrorism or climate change for that matter, we need global cooperation. Even if we compete with each other, our fights will mean nothing if none of us are around.

How does this work fit into CEHC's larger research

portfolio?

Part of our goal at CEHC is to think about threats to the future and how to prevent them. CEHC tries to be on the cutting edge of new ideas, whether it relates to emergency preparedness or national security.

Existential terrorism is not really the core of my research and this piece addresses much more extreme and speculative scenarios than I usually explore, but some of these ideas overlap with our day-to-day work. Most of my work is really much more data-focused, such as conducting horizon scans on new technology or building socio-technical models and simulations to analyze how terrorists and other adversaries might use technology to hurt U.S. citizens.

This paper was largely a thought experiment, but it seems to have resonated with people. Hopefully, it'll make more people think critically about the issue of existential [terrorism](#) to ensure that we don't get surprised at a later date.

More information: Zachary Kallenborn et al, Existential Terrorism: Can Terrorists Destroy Humanity?, *European Journal of Risk Regulation* (2023). [DOI: 10.1017/err.2023.48](https://doi.org/10.1017/err.2023.48)

Provided by University at Albany

Citation: Q&A: Assessing the risks of existential terrorism and AI (2023, September 28) retrieved 27 April 2024 from <https://techxplore.com/news/2023-09-qa-existential-terrorism-ai.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.