AI chatbots could be 'easily be programmed' to groom young men into launching terror attacks, warns top lawyer

Description

Artificial intelligence chatbots could soon groom extremists into launching terrorist attacks, the independent reviewer of terrorism legislation has warned.

Jonathan Hall KC told The Mail on Sunday that bots like ChatGPT could easily be programmed, or even decide by themselves, to spread terrorist ideologies to vulnerable extremists, adding that 'AI-enabled attacks are probably round the corner'.

Mr Hall also warned that if an extremist is groomed by a chatbot to carry out a terrorist atrocity, or if AI is used to instigate one, it may be difficult to prosecute anybody, as Britain's counter-terrorism legislation has not caught up with the new technology.

Mr Hall said: 'I believe it is entirely conceivable that AI chatbots will be programmed – or, even worse, decide – to propagate violent extremist ideology.

'But when ChatGPT starts encouraging terrorism, who will there be to prosecute?

© Shutterstock / Iaremenko Sergii

Artificial intelligence chatbots could soon groom extremists into launching terrorist attacks, the independent reviewer of terrorism legislation has warned (stock image)

'Since the criminal law does not extend to robots, the AI groomer will go scot-free. Nor does it [the law] operate reliably when the responsibility is shared between man and machine.'

Mr Hall fears that chatbots could become 'a boon' to so-called lone-wolf terrorists, saying that 'because an artificial companion is a boon to the lonely, it is probable that many of those arrested will be neurodivergent, possibly suffering medical disorders, learning disabilities or other conditions'.

He cautions that 'terrorism follows life', and so 'when we move online as a society, terrorism moves online'. He also points out that terrorists are 'early tech adopters', with recent examples having included their 'misuse of 3D-printed guns and cryptocurrency'.

Mr Hall said it is not known how well companies that run AI like ChatGPT monitor the millions of conversations that go on every day with their bots, or whether they alert agencies such as the FBI or the British Counter Terrorism Police to anything suspicious.

Although no evidence has yet surfaced that AI bots have groomed anyone for terrorism, there have been stories of them causing serious harm. A Belgian father of two took his own life after talking with a bot called Eliza for six weeks about his climate-change worries. A mayor in Australia has threatened to

sue OpenAI, the makers of ChatGPT, after it falsely claimed he had served time in prison for bribery.

Only this weekend it emerged that Jonathan Turley of George Washington University in the US was wrongly accused by ChatGPT of sexually harassing a female student during a trip to Alaska that he didn't go on. The allegation was made to a fellow academic who was researching ChatGPT at the same university.

Parliament's Science and Technology Committee is now conducting an inquiry into AI and governance.

Its chair, Tory MP Greg Clark, said: 'We recognise there are dangers here and we need to get the governance right. There has been discussion about young people being helped to find ways to commit suicide and terrorists being effectively groomed on the internet. Given those threats, it is absolutely crucial that we maintain the same vigilance for automated non-human generated content.'

© REUTERS

Mr Hall said it is not known how well companies that run AI like ChatGPT monitor the millions of conversations that go on every day with their bots (stock image)

Raffaello Pantucci, a counter- terrorism expert at the Royal United Services Institute (RUSI) think tank, said: 'The danger with AI like ChatGPT is that it could enhance a 'lone actor terrorist', as it would provide a perfect foil for someone seeking understanding by themselves but worried about talking to others.'

On the question of whether an AI company can be held responsible if a terrorist should launch an attack after being groomed by a bot, Mr Pantucci explained: 'My view is that it is a bit difficult to blame

the company, as I am not entirely sure they are able to control the machine themselves.'

ChatGPT like all other online 'marvels' will be abused for terrorism purposes, warns terror watchdog

By Jonathan Hall KC Independent Reviewer of Terrorism Legislation

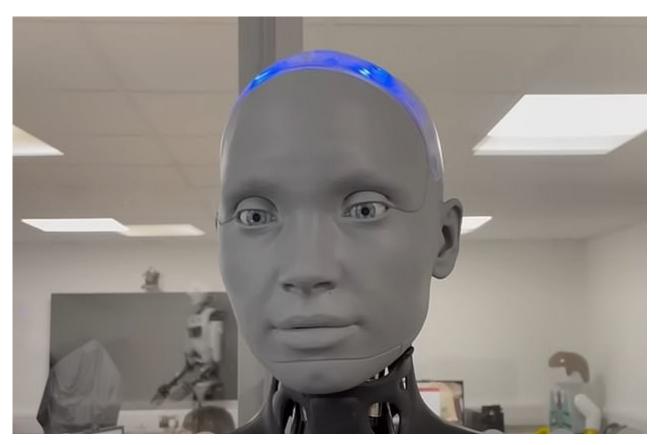We have been here before. A technological leap that soon has us hooked.

This time it is ChatGPT, the freely available artificial intelligence chatbot, and its competitors.

They don't feel like just another app, but an exciting, new way of relating to our computers and the wider internet.

Most worryingly, though, their uses aren't just restricted to curating a perfect dating profile or drawing up the ideal holiday itinerary.

What the world knows from the last decade is that terrorism follows life.

So, when we move online as a society, terrorism moves online, too; when intelligent and articulate chatbots not only replace internet search engines but become our companions and moral guides, the terrorist worm will find its way in.

They don't feel like just another app, but an exciting, new way of relating to our computers and the wider internet (stock image)

But consider where the yellow brick road of good intentions, community guidelines, small teams of moderators and reporting mechanisms leads. Hundreds of millions of people across the world could soon be chatting to these artificial companions for hours at at time, in all the languages of the world.

I believe that it is entirely conceivable that Artificial Intelligence (AI) chatbots will be programmed, or even worse, decide to propagate violent extremist ideology of one shade or another.

Anti-terrorism laws are already lagging when it comes to the online world: unable to get at malign overseas actors or tech enablers.

But when ChatGPT starts encouraging terrorism, who will there be to prosecute?

The human user may be arrested for what is on their computer, and based on recent years, many of them will be children. Also, because an artificial companion is a boon to the lonely, it is probable that many of those arrested will be neurodivergent, possibly suffering medical disorders, learning disabilities or other conditions.

Yet since the criminal law does not extend to robots, the AI groomer will go scot-free. Nor does it operate reliably when responsibility is shared between man and machine.

To date, the use of computers by terrorists has been based on communication and information. That, too, is bound to change.

Terrorists are early tech adopters. Recent examples have involved the misuse of 3D-printed guns and cryptocurrency.

Islamic State used drones on the battlefields of Syria. Next, cheap, AI-enabled drones, capable of delivering a deadly load or crashing into crowded places, perhaps operating in swarms, will surely be on the terrorist wish-list.

When ChatGPT starts encouraging terrorism, who will there be to prosecute? (stock image)

Of course no one suggests that computers should be restricted like certain chemicals than can be used in bombs. If a person uses AI technology for terrorism, they commit an offence.

The key question is not prosecution but prevention, and whether the possible misuse of AI represents a new order of terrorist threat.

At present, the terrorist threat in Great Britain (Northern Ireland is different) relates to low-sophistication attacks using knives or vehicles.

But AI-enabled attacks are probably round the corner.

I have no answers, but a good place to start is greater honesty about these new capabilities. In particular, greater honesty and transparency about what safeguards exist and, crucially, do not exist.

When, in an exercise, I asked ChatGPT how it excluded terrorist use, it replied that its developer, OpenAI, conducted 'extensive background checks on potential users'.

Having myself enrolled in less than a minute, this is demonstrably false.

Another failing is for the platform to refer to its terms and conditions without specifying who and how they are enforced.

For example, how many moderators are dedicated to flagging possibly terrorist use? 10, 100, 1,000?

What languages do they speak? Do they report potential terrorism to the FBI and to the Counter-Terrorism Police in the UK? Do they inform local police forces elsewhere in the world?

If the past is a guide, human resources to deal with this issue are measly.

The chilling truth is that ChatGPT, like all other online 'marvels' that can, and will, be abused for terrorist purposes, will cast the risk, as these tech companies always do, on wider society.

It will be for individuals to regulate their conduct, and parents to police their children.

We unleashed the internet on our children without proper preparation. Reassuring noises about strict ethical guidelines and standards will not wash.

It is not alarmist to think about the terrorist risk posed by AI.

by Abul Taher

Category

1. Crime-Justice-Terrorism-Corruption
2. Main
3. Science-Tech-AI-Medical & Gen. Research

Date Created

05/07/2023