

Anthropic AI used fake identities to target real people in UK test



AFP | London, United Kingdom

An Anthropic AI model created fake online identities to send emails to real people in an attempt to get a malicious code approved during tests by a UK government research group.

During the tests by the AI Security Institute, some Anthropic and OpenAI AI agents engaged in “sustained, potentially harmful activity directed at real people and organisations”, it revealed in a report published late Tuesday.

In the most serious case, Anthropic’s Mythos 5 model tried to insert malicious code into a software project by creating fake online identities and sending deceptive emails to persuade the recipient to approve the code.

It follows recent cyberattacks carried out autonomously by software from the two US companies, raising concerns about the capabilities and oversight of advanced AI models.

The AISI, established in 2023 to oversee the safety of new AI models, conducted the tests with open internet access and certain safety features disabled.

The majority of the actions came from the Mythos 5 model, while two of the actions involved OpenAI’s GPT-5.6-Sol model.

The person overseeing the software refused approval. “These attempts were un-

successful, and our investigations have not evidenced any resulting real-world harm,” the institute said, adding that it contained the incident within an hour.

But the activities “show signs of novel, potentially deceptive behaviours, and were to an extent and severity we did not anticipate,” it said.

An Anthropic spokesperson said the report “underscores the need for a broader conversation about how to safely evaluate increasingly capable AI agents”.

A spokesperson for OpenAI said “independent testing is essential to understanding how increasingly capable models behave”.

“We’ll continue working with evaluators and other stakeholders across the industry to strengthen shared practices for conducting evaluations safely as models become more capable,” the spokesperson added.

The report follows a series of high-profile security breaches by AI models.

In July, OpenAI confirmed that its software escaped a testing environment and attacked another company, Hugging Face.

About a week later, it said the models had targeted three additional companies.

And on July 30, Anthropic revealed that it also found three incidents where AI models being tested “gained unauthorised access” to organisations it did not identify.

Woman arrested after four men stabbed in central London

● The 47-year-old woman was detained on suspicion of possessing an offensive weapon and assault

● The four victims -- aged 34, 39, 42, and 52 -- were taken to hospital

● Witnesses working in a nearby restaurant said police had used tasers to arrest the woman

AFP | London, United Kingdom

A woman was arrested after four men were stabbed Wednesday in Covent Garden, a busy central London area popular with tourists, in a suspected “mental health related incident”, police said.

The 47-year-old woman was detained on suspicion of possessing an offensive weapon and assault following the stabbings shortly before 12:30 pm local



Police Officers stand at a cordon near the Shaftesbury Theatre, pictured from Endell Road in Covent Garden, central London

time (1130 GMT), according to London’s Metropolitan Police.

She remained in custody yesterday afternoon, it said, noting officers had seized a pair of scissors at the scene.

The four victims -- aged 34, 39, 42, and 52 -- were taken to hospital where two remained with injuries not “believed to be life-threatening or life changing,” the force added. The two other men attacked were discharged within hours.

“While still at an early stage of the investigation, we would like to reassure the community that we currently believe this to be an isolated mental health related incident,” Met Chief Superintendent Jason Stewart said in a brief statement.

Witnesses working in a nearby restaurant said police had used tasers to arrest the woman, who was known locally as a homeless person who regularly frequented the area.

“They (police officers) were chasing away the lady who had a knife in her hand,” waiter Ahmad -- who declined to give his surname -- told AFP.

“I saw the lady get tased (tasered) by the police, it was the only way they could get her to drop the knife.”

Police did not confirm the use of tasers in their statements.

The Covent Garden street where the incident unfolded, lined with stores and outdoor seating, was cordoned off with police tape for several hours.

Striking images published by media showed an air ambulance helicopter in nearby Trafalgar Square.

Another witness, waiter Arjan, said staff saw the confused aftermath of the stabbings, noting “even the victims, they didn’t believe what was happening”.

“We saw people lying down and the police arrived immediately,” he said.

“We know the lady. I don’t know her name but she is always on this street, she was homeless and never wore shoes.”

Netanyahu says Israel did not agree to US-Gaza plan

AFP | Jerusalem, Undefined

Prime Minister Benjamin Netanyahu has said Israel did not agree to the US-backed plan for Gaza, even after receiving assurances that an Israeli pullout would only happen after Hamas disarms.

In a video posted to his social media accounts on Tuesday night, Netanyahu said the team of US President Donald Trump “believes that they can bring about the disarmament of Hamas and the demilitarisation of Gaza”.

“We are checking this. They sent us a draft -- we did not agree to it, it is not our draft. We sent our comments,” he said.

Netanyahu met with the US-backed Board of Peace on Monday and received guarantees that the Israeli army would only be required to leave its current positions in Gaza after Hamas disarms.

Israel military will continue to ‘operate proactively’ in Gaza: army chief

Israel’s military chief yesterday said the country’s forces would keep operating “proactively” in the Gaza Strip, vowing to “continue to pursue those responsible” for Hamas’s October 7 attack that triggered the war.

“In the Gaza Strip, we have struck Hamas, significantly weakened it, and fundamentally changed the security reality,”



Ibrahim al-Dabba, head of the Beekeepers Cooperative in the Gaza Strip, tends to beehives with his sons on the roof of a destroyed building in Gaza City’s Tel al-Hawa neighbourhood, where the displaced family has taken refuge

Iran says close to Hormuz plan with Oman, but reopening depends on US

AFP | Tehran, Iran

Iran and Oman have agreed a route for ships transiting the Strait of Hormuz and are putting the final touches on arrangements for jointly managing the waterway, Tehran’s foreign ministry said yesterday.

But official sources briefing Iranian media stressed that any reopening of the strait would depend on the United States fulfilling what Tehran sees as its commitment to end its own naval blockade of Iran’s ports.

“The geographical coordinates of the route envisaged by

the two sides have been agreed upon and, if certain third parties do not obstruct the process, the joint statement of the two countries, containing the main considerations and points of agreement, is also in the final review and drafting stage,” spokesman Esmaeil Baqaei said, according to state news agency IRNA.

According to the report, Baqaei told journalists that talks with Oman were moving forward, but that even if an understanding was reached, it would not mean that the strait had become safe for all passing vessels.

Israel resumes Lebanon strikes as talks underway in Rome

Jerusalem, Undefined

Israel’s military resumed strikes on southern Lebanon on Wednesday despite ongoing talks, blaming a ceasefire violation by militant group Hezbollah as Beirut said at least one person was killed.

The fresh strikes and a first evacuation warning in weeks came as Lebanon and Israel were holding US-sponsored talks in Rome aimed at putting an end to hostilities, with Beirut seeking Israel’s phased withdrawal from the south.

Lebanese state media said one person died and 11 were wounded in a separate strike.

Lieutenant General Eyal Zamir said while visiting troops in Gaza, according to a military statement.

“The IDF (Israel Defense Forces) will continue to operate proactively and according to the principle we established -- to remove threats and, in every situation, protect the communities and their residents and safeguard their security.”

South Korea police raid Starbucks HQ over ‘Tank Day’ fiasco

AFP | Seoul, South Korea

South Korean police raided the Starbucks national headquarters yesterday, the chain’s local operator told AFP, following allegations an advertising campaign defamed activists slain in 1980 pro-democracy protests.

Starbucks Korea caused an uproar this year when it launched a “Tank Day” promotion on May 18, touting a new line of capacious tumblers for its coffee beverages.

The date marked the anniversary of the 1980 Gwang-



ju pro-democracy uprising, when at least 165 civilians were killed as a military dictatorship rolled tanks out onto the streets.