North Korea and Iran using AI for hacking, Microsoft says

Publication Date: 2024-02-14

Author: Unknown

Section: Technology

Tags: Hacking, Microsoft, OpenAI, Artificial intelligence (AI), North Korea, US foreign policy, Iran, news

Article URL: https://www.theguardian.com/technology/2024/feb/14/north-korea-iran-ai-hacking-microsoft

US adversaries – chiefly Iran and North Korea, and to a lesser extent Russia and China – are beginning to use generative artificial intelligence to mount or organize offensive cyber operations, Microsoft said on Wednesday. Microsoft said it detected and disrupted, in collaboration with business partner OpenAI, many threats that used or attempted to exploit AI technology they had developed. In a blogpost, the company said the techniques were “early-stage” and neither “particularly novel or unique” but that it was important to expose them publicly as US rivals leveraging large-language models to expand their ability to breach networks and conduct influence operations. Cybersecurity firms have long used machine-learning on defense, principally to detect anomalous behavior in networks. But criminals and offensive hackers use it as well, and the introduction of large-language models led by OpenAI’s ChatGPT upped that game of cat-and-mouse. Microsoft has invested billions of dollars in OpenAI, and Wednesday’s announcement coincided with its release of a report noting that generative AI is expected to enhance malicious social engineering, leading to more sophisticated deepfakes and voice cloning. A threat to democracy in a year where over 50 countries will conduct elections, magnifying disinformation and already occurring, Microsoft provided some examples. In each case it said all generative AI accounts and assets of the named groups were disabled: • The North Korean cyber-espionage group known as Kimsuky has used the models to research foreign thinktanks that study the country, and to generate content likely to be used in spear-phishing hacking campaigns. • Iran’s Revolutionary Guard has used large-language models to assist in social engineering, in troubleshooting software errors and even in studying how intruders might evade detection in a compromised network. That includes generating phishing emails “including one pretending to come from an international development agency and another attempting to lure prominent feminists to an attacker-built website on feminism”. The AI helps accelerate and boost the email production. • The Russian GRU military intelligence unit known as Fancy Bear has used the models to research satellite and radar technologies that may relate to the war in Ukraine. • The Chinese cyber-espionage group known as Aquatic Panda – which targets a broad range of industries, higher education and governments from France to Malaysia – has interacted with the models “in ways that suggest a limited exploration of how LLMs can augment their technical operations”. • The Chinese group Maverick Panda, which has targeted US defense contractors among other sectors for more than a decade, had interactions with large-language models suggesting it was evaluating their effectiveness as a source of information “on potentially sensitive topics, high profile individuals, regional geopolitics, US influence, and internal affairs”. In a separate blog published on Wednesday, OpenAI said its current GPT-4 model chatbot offers “only limited, incremental capabilities for malicious cybersecurity tasks beyond what is already achievable with publicly available, non-AI powered tools”. Cybersecurity researchers expect that to change. Last April, the director of the US Cybersecurity and Infrastructure Security Agency, Jen Easterly, told Congress that “there are two epoch-defining threats and challenges. One is China, and the other is artificial intelligence.” Easterly said at the time that the US needed to ensure AI is built with security in mind. Critics of the public release of ChatGPT in November 2022 – and subsequent releases by competitors including Google and Meta – contend it was irresponsibly hasty, considering security was largely an afterthought in their development. “Of course bad actors are using large-language models – that decision was made when Pandora’s Box was opened,” said Amit Yoran, chief executive of the cybersecurity firm Tenable. Some cybersecurity professionals complain about Microsoft’s creation and hawking of tools to address vulnerabilities in large-language models when it might more responsibly focus on making them more secure. “Why not create more secure black-box LLM foundation models instead of selling defensive tools for a problem they are helping to create?” asked Gary McGraw, a computer security veteran and co-founder of the Berryville Institute of Machine Learning. The NYU professor and former AT&T chief security officer Edward Amoroso said that while the use of AI and large-language models may not pose an immediately obvious threat, they “will eventually become one of the most powerful weapons in every nation-state military’s offense”.