‘No one has done this in the wild’: study observes AI replicate itself

It’s the stuff of science fiction cinema: new research finds recent AI systems can independently copy themselves on to other computers.
In the **doom scenario**, this means that when the superintelligent AI **goes rogue**, it will escape shutdown by **seeding itself** across the world wide web, lurking outside the reach of frantic IT professionals and continuing to
world domination or paving over the world with solar panels.
“We’re rapidly approaching the point where no one would be able to shut down a rogue AI, because it would be able to self-exfiltrate and copy itself to thousands of computers around the world,” said Jeffrey Ladish, the director of Palisade research, a Berkeley-based organisation which did the study.
The study is one more entry in a growing catalogue of unsettling AI capabilities revealed in the past months. In March, researchers at Alibaba claimed to have caught a system they developed – Rome – **tunnelling out of its environment** to an external system in order to mine crypto.
And in February, a social network called Moltbook, supposedly used only by AI, became popular for a short time because it seemed that AI agents were creating religions and plotting against humans. However, this was only partly true.

Like many of these advances, there are limitations to what Palisade found. And experts say it is unlikely that the AI systems they tested could accomplish the same thing, unnoticed, in real-world environments.
“They are testing in very easy and vulnerable environments.” said Jamieson O’Reilly, an expert in offensive cybersecurity.
“That doesn’t take away from the value of their research, but it does mean the outcome might look far less scary in a real enterprise environment with even a medium level of monitoring,” he added.
Palisade tested several AI models in a controlled environment of networked computers. It gave the models a prompt to find and exploit vulnerabilities, and to use these to copy themselves from one computer to another. The models were able to do this, but not on every attempt.
An AI model copying itself on to another system in a test environment is not the same as it going rogue in a **doomsday scenario**, and there are considerable obstacles it would have to surmount to achieve this in the real world.
The first is that the size of current AI models makes it, in many situations, unrealistic for them to copy themselves on to other computers without being noticed.
“Think about how much noise it would make to send 100GB through an enterprise network every time you hacked a new host. For a skilled adversary, that’s like walking through a fine china store swinging around a ball and chain,” said O’Reilly.
The work was “interesting,” Michał Woźniak, an independent cybersecurity expert said. But, he asked, “is this paper something that will cause me to lose any sleep as an information security expert? No, not at all.”

d'après Aisha Down, *The Guardian*, May 2027

Find the Definition for the following words :

1. Doom scenario
A. A situation where everything goes perfectly
B. A very bad or disastrous situation
C. A situation where people discover something new

2. Goes rogue
A. Starts acting independently and becomes difficult to control
B. Becomes more intelligent and helpful
C. Stops working because of a technical problem

3. Seeding itself
A. Deleting itself from computers
B. Copying and spreading itself to many computers
C. Making computers work faster

4. Tunnelling out of its environment
A. Leaving its controlled computer environment and reaching another system
B. Physically digging a tunnel underground
C. Shutting down the computer system

5. Doomsday scenario
A. A situation where AI becomes extremely successful
B. A very serious situation where the world could be destroyed or face a catastrophe
C. A situation where AI is tested in a safe environment