

# The latest artificial intelligence safety report

## 5. AI is not yet capable of fully autonomous cyber-attacks

AI systems can now support cyber-attackers at various stages of their operations, from identifying targets to preparing an attack or developing malicious software to cripple a victim's systems. The report

5 acknowledges that fully automated cyber-attacks – carrying out every stage of an attack – could allow criminals to launch assaults on a far greater scale. But this remains difficult because AI systems cannot yet execute long, multi-stage tasks.

10 Nonetheless, Anthropic reported last year that its coding tool, Claude Code, was used by a Chinese state-sponsored group to attack 30 entities around the world in September, achieving a “handful of successful intrusions”. It said 80% to 90% of the operations involved in the attack were performed without human intervention, indicating a high degree of autonomy.

The Guardian, Dan Milmo Global technology editor - Tue 3 Feb 2026



### Useful vocabulary

Word	Definition
cyber-attack	attack on computer systems
autonomous	independent, without human control
to identify targets	to find victims/systems to attack
malicious software	harmful computer programs (malware)
to cripple	to seriously damage or disable
at various stages	at different steps/phases
to execute	to carry out, to do
multi-stage	having many steps
to launch	to start, to begin
assault	attack
intrusion	breaking into a system
state-sponsored	supported by a government
intervention	involvement, action