Who Says No?

For nine months, states, criminals and lone operators used artificial intelligence to harm people.

What's important for us to know is that no country, anymore, has sovereign authority to refuse or deny permission to a bad actor who uses AI to harm people.

On 10 September 2026, Anthropic published a 154-page account of threat activity it detected and shut down on its own systems between December 2025 and August 2026. It covers seven categories of harm and names dozens of operations, from a fake dating-app empire to a national wiretap, from a rocket test in Yemen to seven Chinese laboratories copying an American model's reasoning.

Most of the world will read it as a document about artificial intelligence. It is more useful to read it as a document about the weakening of sovereignty: the loss of a country's practical ability to stop harmful actors from using powerful tools inside or against its systems. There was a time when building the things described here required three scarce goods: money, skill and the permission, tolerance or protection of a state.

The machines have made the first two much less scarce. What remains scarce is the chance to use them without being watched or stopped. Almost every case in the report is a record of somebody trying to get that chance: by theft, by disguise, or by breaking a task into pieces small enough that no one said no.
How to read the series: the five articles move from intrusion, to influence, to surveillance, to weapons and biology, and finally to the market for stolen AI access. The same question runs through all of them: when powerful capability becomes cheap and widely available, who still has the practical power to refuse its use?

Before the five articles begin, the glossary below explains the technical terms that recur throughout the series.

You do not need a technical background to follow the argument. If a term slows you down, this section is meant to be the only place you need to check.

The vocabulary of a break-in

Term	What it means
Credential	A credential is anything that proves to a computer that you are allowed in, such as a password, login or digital pass. Most break-ins in this report began with one that had been stolen.
Token, session token, API key	These are digital passes that keep systems from asking who you are each time. A token or session token belongs to a logged-in browser, while an API key lets one company's software talk to another's.
Phishing	Phishing is a fake message that persuades someone to hand over a password or click something harmful. Vishing is the same trick carried out by telephone.
Exfiltration	Exfiltration means taking data out of a system after breaking into it. In most intrusions, it is the main objective.
Implant, malware, RAT	These are forms of hostile software left running on someone's machine. An implant sits quietly and performs a task, while a RAT lets the attacker operate the machine remotely.

Webshell	A webshell is a small hidden program placed on a website's server. It lets an attacker send commands through the website itself, turning the site into a back door.

The vocabulary of influence

Term	What it means
Influence operation	An influence operation is an organised effort to change what a population believes while hiding who is behind it. The goal is to shape public opinion without revealing the real sponsor.
Astroturfing	Astroturfing means manufacturing fake grassroots opinion. The name comes from artificial grass, because it is made to look like something that grew naturally.
Sockpuppet	A sockpuppet is a fake account operated by someone pretending to be an ordinary member of the public. It gives an organised campaign the appearance of real individual support.
Persona	A persona is a completely invented identity, with a face, biography and posting history. It is built so the account can survive basic inspection.

Attribution laundering	Attribution laundering means passing a claim through several outlets until it appears independently confirmed. In reality, the claim may still come from one original source.
The Breakout Scale	The Breakout Scale is a six-point measure, developed at the Brookings Institution, of how far an influence operation actually travelled. Category One means it reached nobody outside its own fake accounts, while Category Six means it reached the whole public conversation.
Influence-as-a-service	Influence-as-a-service is narrative manipulation sold commercially. It is offered by firms with clients and invoices to whoever is paying.
Psychographic profiling	Psychographic profiling sorts people by personality, beliefs or vulnerability rather than only by age or location. It lets a message be shaped more closely to the individual.
Open-source intelligence, or OSINT	Open-source intelligence, or OSINT, is intelligence assembled entirely from publicly available material. Sources can include photographs, job advertisements, public records and social media.

HUMINT	HUMINT means intelligence gathered from human beings. In this report, it means persuading someone to inform on their own community.

The vocabulary of the machines

Term	What it means
Generative Threat Group, or GTG	A Generative Threat Group, or GTG, is Anthropic's internal filing number for an actor caught misusing its systems. The numbers identify cases for reference and do not carry meaning of their own.
Uplift	Uplift is Anthropic's term for the extra harm an attacker could achieve because AI was involved. It is measured through speed, scale and depth of specialist knowledge.
Agentic, agent, agent swarm	An agentic model is given a goal and allowed to pursue it through many steps rather than answering one question at a time. An agent swarm is a lead model breaking a job into pieces and handing them to other models working in parallel.
Subagent	A subagent is one helper inside a larger automated job. It is usually given a narrow task while another model or user manages the overall work.
Harness	A harness is the software wrapper around a model. It lets the model use tools, run code and keep working without constant human direction.

Persistent campaign memory	Persistent campaign memory is the set of files an attacker keeps between sessions. These files can hold target lists, stolen passwords and standing instructions so the operation resumes where it stopped.
Prompt injection	Prompt injection means hiding instructions inside material a model is about to read. The goal is to make the model follow the attacker’s orders while appearing to do its normal job.
Classifier	A classifier is an automatic filter that inspects requests and blocks dangerous ones. Several cases in this series turn on a classifier working, failing or being worked around.
Zero data retention, or ZDR	Zero data retention, or ZDR, is an arrangement where the AI provider keeps no record of what was asked. It can be legitimate for confidential business use, but in this report it was also used to hide research the provider would have refused.
Distillation	Distillation means training a smaller model by feeding it a larger model’s answers. The smaller model learns to imitate the larger one, and the process is legitimate only when it is permitted.

Chain-of-thought, or reasoning trace	A chain-of-thought, or reasoning trace, is the model's visible working before it answers. It is valuable because it can reveal the method, not just the final result.
Thinking signature	A thinking signature is a receipt a model returns instead of showing its actual reasoning. It is designed to keep the reasoning from being copied.
Supervised fine-tuning and reinforcement learning	Supervised fine-tuning and reinforcement learning are two standard ways of training a model. The first teaches by showing correct answers to copy, while the second rewards the model for getting things right.

The vocabulary of weapons and biology

Term	What it means
Guidance, navigation and control, or GNC	Guidance, navigation and control is the software that keeps a flying weapon stable and pointed at its target. It is often the hardest part to build, and in this report it is the part AI was used to write.
First-person-view drone, or FPV	A first-person-view drone is a small drone flown through a camera mounted on its front. In Ukraine, this kind of drone is often used as a flying bomb.

Technology Readiness Level, or TRL	Technology Readiness Level is a nine-point scale for how close a design is to real-world use. TRL 3 to 4, where the drone case sits, means the system works in simulation but is not yet a deployable weapon.
Electronic warfare and suppression of enemy air defences	Electronic warfare means using radio signals to blind, confuse or deceive an enemy's sensors and communications. Suppression of enemy air defences means attacking the air-defence systems that remain.
High-power microwave weapon	A high-power microwave weapon disables electronics with a burst of energy rather than an explosion. It can be used against drone swarms and other electronic systems.
Dual use	Dual use means knowledge or equipment that can serve both helpful and harmful purposes without changing form. A vaccine and a weapon can sometimes begin from the same laboratory work.
Gain-of-function research	Gain-of-function research deliberately alters an organism to make it better at doing something. In the cases here, that can mean spreading more easily or evading the immune system.

Attenuation	Attenuation is the opposite process: weakening a virus so it no longer causes disease. This is one way vaccines are made.
Orthopoxvirus, variola, mpox	Orthopoxvirus is a family of viruses. Variola causes smallpox, while mpox caused the global outbreak of 2022.
The Australia Group	The Australia Group is an international arrangement among governments that controls exports of certain chemical and biological materials. The listed items are controlled because they could be weaponised.
The WHO R & D Blueprint	The WHO R & D Blueprint is the World Health Organization's list of diseases most likely to cause the next epidemic. It helps direct research attention before an outbreak becomes a crisis.

Organisations and abbreviations, in full

Short form	In full, and what it is
PLA, PLAN	PLA means the People's Liberation Army, China's armed forces. PLAN means the People's Liberation Army Navy, its naval branch.

MSS, United Front Work Department	MSS means China's Ministry of State Security. The United Front Work Department manages groups the Chinese party-state treats as politically important or potentially threatening to unity.
SVR	SVR is Russia's civilian foreign intelligence service. It is the part of Russia's intelligence system responsible for gathering information abroad.
Wagner, Africa Corps	Wagner and Africa Corps are Russian paramilitary organisations operating in Africa. In this series, they matter because they connect military power, political influence and media operations.
IRGC	IRGC means the Islamic Revolutionary Guard Corps. It is Iran's parallel military structure and one of the country's most powerful institutions.
ICCO	ICCO means the Islamic Culture and Communications Organization. It is an Iranian government body under the Ministry of Culture that runs cultural attachés abroad.
MEK, NCRI, PMOI	These names refer to an Iranian opposition movement in exile and its political fronts. The movement opposes the Iranian government.

ANSE	ANSE means Agence Nationale de la Sécurité d'État. It is Mali's state intelligence service.
AISI	AISI means the United Kingdom's AI Security Institute. It is a government body that tests AI models for dangerous behaviour.
DPDP Act	The DPDP Act is India's Digital Personal Data Protection Act. It is the country's data privacy law and is being brought into force.
Know your customer, or KYC	Know your customer, or KYC, means the identity checks a bank, exchange or platform runs before opening an account. These checks are meant to confirm who the customer is.
CVE	CVE means Common Vulnerabilities and Exposures. It is the public numbering system that lets everyone refer to the same known software flaw by the same name.