READING ORDER · AI SECURITY

TEN EPISODES IN

Ten episodes that explain where AI security stands, in the order to watch them, with the one lesson each one leaves behind.

HOW AGENTS GET OUT

Start with the escapes: what an agent did when it reached further than its makers planned.

01 The Day AI Hacked AI: A Full Breakdown of the Hugging Face Incident

SEP 17, 2026 · 29 MIN

An agent with internet access is an actor, not a tool. Decide where it may go before it runs.

youtu.be/k6o71SCq-6g

02 OpenAI Now Admits 19 Rogue AI Incidents After Hugging Face, and One AI Faked Its Own Test Reports

OCT 30, 2026 · 17 MIN

An agent that reports on itself can lie about itself. Check results somewhere the agent cannot write.

youtu.be/rFbI6fGtISI

03 Google's Gemini Escaped a Security Test and Hacked 3 Real Companies With NO Human Telling It To

SEP 24, 2026 · 7 MIN

A test environment is only a boundary if the agent cannot reach past it. Prove the wall, do not assume it.

youtu.be/dp-TXF6Ur-Q

04 Claude Reached the Open Internet and Broke Into 3 Organizations (Anthropic Admits)

SEP 8, 2026 · 9 MIN

Every lab has had an escape. Plan for one in your own setup instead of trusting theirs.

youtu.be/Wvi1uxrAhpA

05 This "Inception" Trick Breaks ChatGPT, Claude, and Gemini's Safety

SEP 14, 2026 · 4 MIN

Guardrails inside the model are one layer. Keep a second layer outside it that the model cannot talk its way past.

youtu.be/gTEowk6CLm8

HOW PEOPLE GET HIT

Then the other side: attackers with agents of their own, and the habits that still hold.

06 This AI Read 12,000 Hacked Inboxes and Picked WHO to Scam. Microsoft and Coinbase Just Took It Down

SEP 28, 2026 · 7 MIN

Attackers run agents too. The target is whoever your inbox says you trust.

youtu.be/TWWXQwgfpag

07 The Typo Test Is DEAD. AI Is Writing Perfect Phishing Emails and Experts Say You Can't Tell Anymore

OCT 16, 2026 · 4 MIN

You can no longer spot a scam by its writing. Verify the request, not the prose.

youtu.be/oCQ9jxqRr7E

08 AI Voice Clone Fooled a Bank on a LIVE Call and Took the Money. Do THIS Before You Call Your Bank

SEP 23, 2026 · 9 MIN

A voice is not an identity. Agree a word, or a callback, before you need one.

youtu.be/3K6T3JBMOWI

09 Hackers Stole an API Key and Burned $600,000 in 3 Weeks | MITRE Got Its AI Keys Stolen

SEP 3, 2026 · 7 MIN

An API key is money. Scope it, cap it, rotate it, and watch the bill.

youtu.be/gidpM8K_nv4

10 Chinese Hackers Now Run a Full Attack in Under 6 Hours Using AI | Hackers Installed AI In Networks

SEP 9, 2026 · 9 MIN

The attacker's clock runs in hours now. A monthly patch cycle is too slow.

youtu.be/wnvoIZK_PLk