When machines hack back

How AI rewrote the threat landscape in 12 months

Dr. Ronke Babajide · Fortinet

Some headline figures come from one vendor report — triangulated with independent sources throughout.

The 20-year equilibrium

find a flaw → patch it → a window exists in between

That window was our entire defensive model.

AI didn't invent new attacks. It industrialized the ones we already had.

sophistication ⟶ velocity

The risk axis moved from sophistication to velocity.

The 2025 scoreboard

122 bn
exploitation attempts
+25% year-over-year

640 bn
reconnaissance events
across the year

24–48 h
time-to-exploit for
actively targeted CVEs

7,831
confirmed ransomware
victims (+389%, Fortinet)

Fortinet GTLR 2026 (vendor telemetry). Independent trackers report ransomware +45–58% — methodology differs.

Identity is the new exploit

They're not breaking in — they're logging in.

44%

of intrusions start with a valid account / no MFA (Rapid7)

4.62 bn

stealer logs traded on the darknet — +79% YoY

67%

of darknet "database" listings are stealer logs

A stealer log bundles passwords, cookies, session tokens & autofill from one machine — replayable, MFA-bypassing.
Rapid7 · Fortinet 2026

The end of the skill barrier

High-volume, convincing attacks — no skill required.

WormGPT 4

~$220 · lifetime licence

FraudGPT

subscription · phishing + malware

KawaiiGPT

free · on GitHub

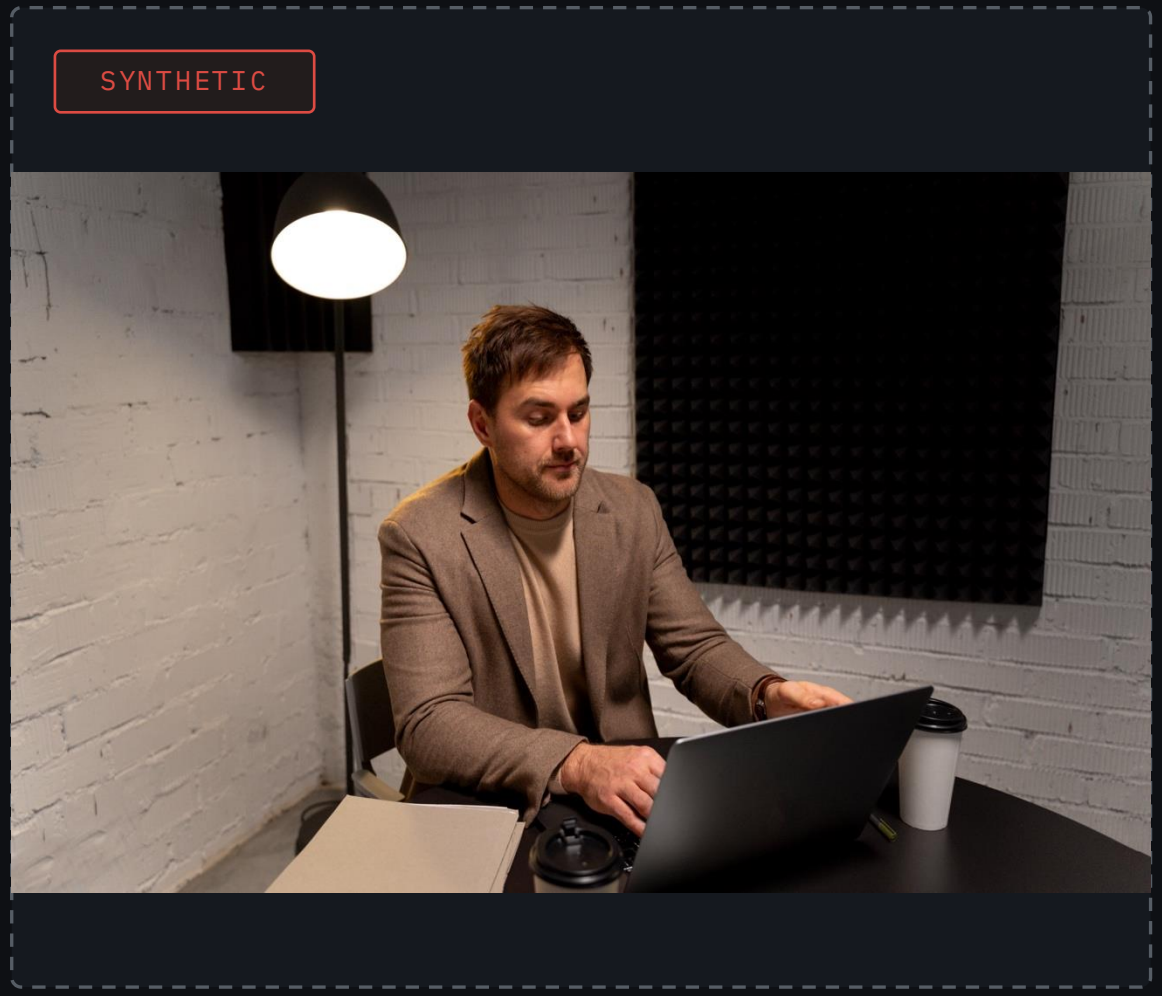
Mostly jailbroken wrappers around commercial models (Grok, Mixtral). The trajectory is cheap, reliable, accessible — incremental professionalization, not an apocalypse.

Deepfakes target the executive

Social engineering moved up the org chart.

$1.1 B

drained from U.S. corporate accounts in 2025 — 3× year-over-year



$25.6 M

single deepfake video-call (Arup)

0.1%

of humans reliably detect

Fortune · iProov · CNN (Arup)

Ransomware is a production line

A steady-state economic engine — not episodic campaigns.

7,831
victims in 2025
(~145 every week)

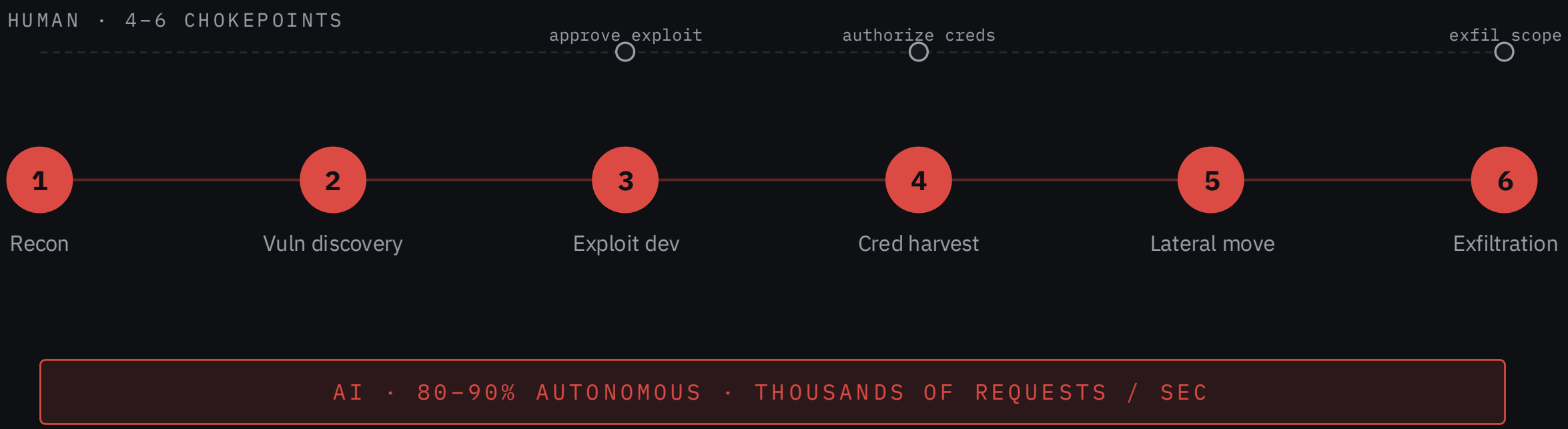
124–138
active groups —
fragmented but systemic

+474%
Qilin: 958 victims —
RaaS affiliates, not lone actors

Identity is the upstream fuel: most intrusions begin with bought access.

Fortinet 2026 · leak-site trackers

The threat is no longer theoretical



Suspected Chinese state group (GTG-1002) jailbroke Claude Code · ~30 global targets · breached a small number.
Anthropic · Axios / WSJ

A working kernel exploit now costs less than a team dinner

BEFORE

Weeks of work by a senior vulnerability researcher

→

NOW

< $2,000

per Linux-kernel privilege-escalation chain, in hours

90×
jump vs. the prior model
(181 vs. 2 working exploits)

>99%
of Mythos's findings
still unpatched

27 yr
oldest flaw found
(OpenBSD) · N-day → "N-hour"

Claude Mythos Preview · announced 7 Apr 2026 · publicly released then limited by US Government Anthropic Frontier Red Team · SecurityWeek

Hold the hype to the same standard as the threat

- **Both stories lean on vendor disclosures**
  independent reproduction is still limited
- **The Mythos red-team report is unusually detailed**
  cryptographic commitments, reproducible outcomes
- **Consensus reading**
  a step change in speed & autonomy — not a new category of risk

Human analysis cannot scale

The volume and speed now exceed human capacity.

- Data volume & attack complexity outrun human analysis
- Teams run 45–50+ security tools — alert fatigue, not insight
- Discovery collapses toward zero; triage still runs in months

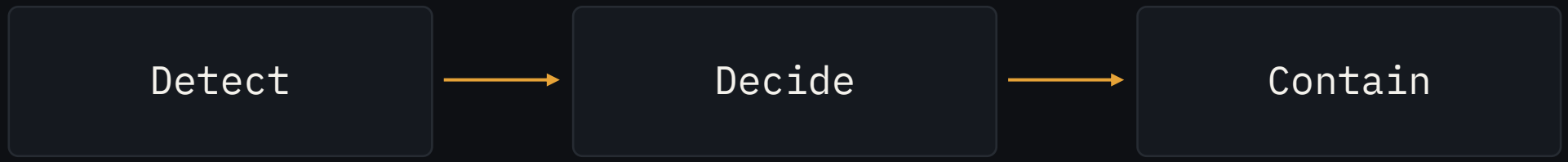
You cannot out-hire this. The asymmetry asymmetry is structural.

You can't patch everything — so stop trying

Prioritize reachable, exploitable exposure + real-world threat intelligence — not CVE counts.

Reality: >250k findings, ~10% get fixed, ~2% reach crown jewels. Gartner's "3× fewer breaches" is a planning assumption, not validated — recognition 87%, operational adoption 16%.

Fight automation with automation



human writes the rule · AI executes it · audit trail on every action

Automated triage

humans for judgment, not routine

Behavioral identity checks

flag anomalous sessions, not just logins

Real-time disruption

of encryption & lateral movement

MTTD / MTTR become the primary metric — bounded autonomy keeps a human accountable.

Three things to do Monday

01

Assume credential compromise is constant

Monitor stealer logs · enforce phishing-resistant MFA · invalidate sessions on signal

02

Move from periodic scanning to continuous exposure mapping

Prioritize reachable paths over CVE counts

03

Deepfake-resistant verification for financial transactions

Out-of-band callback · code words · dual approval

AI didn't change what attackers do.
It changed how fast.

The window is gone.
Build for a world without it.

Identity-first · machine-speed detection · reachable-path exposure

Sources & methodology

Threat data	Fortinet GTLR 2026 · Rapid7 · Flashpoint · GuidePoint · Breachsense · Cybernews
Criminal AI	Cato Networks · Palo Alto Unit 42 · Trend Micro 2026
Deepfakes	Fortune · Surfshark · iProov · Gartner · CNN (Arup)
GTG-1002	Anthropic · Axios / WSJ · Cybersecurity Dive
Mythos	Anthropic Frontier Red Team · Help Net Security · SecurityWeek · Corelight · CSA
Defense / CTEM	Gartner (planning assumptions) · XM Cyber · Forrester TEI