

AI SECURITY

The Fork in the Road

Fable 5 shipped. The exploit window
went negative. Glasswing covers 200
companies. Do the math.



What Shipped on June 9

FABLE 5

First Mythos-class model for everyone. State of the art on nearly every benchmark. Half the price of Mythos Preview.

MYTHOS 5

Same model, safeguards lifted. Cyberdefenders only, via Project Glasswing.

Stripe: a 2 month migration of 50 million lines of Ruby, done in 1 day

The Slowdown Ask



June 4

Anthropic: the world may need the option to pause frontier AI



June 9

Anthropic ships its most capable model ever

Claude already writes 80% of the code in Anthropic's own codebase

Directionally right. Strategically free. The ask costs nothing. The release pays the bills.

The Exploit Window Went Negative



Mean time to exploit, Google M-Trends 2026.
Exploitation now happens a week before a patch exists.

Disclosure to first exploitation: 771 days in 2018. Single digit hours in 2024.

Discovery Is Free. Patching Is Not.

Under 1%

of AI-discovered vulnerabilities
have been patched



530 high or critical bugs disclosed,
75 patched



2 weeks average to patch one critical bug



Open source maintainers asked
Anthropic to slow down its disclosures

**The company that asked the industry to
slow down is being asked to slow down.**

Glasswing Does Not Cover You



200 organizations protected

Hundreds of millions of companies are not

10,000+ high or critical vulnerabilities found

Discovery funding outweighs remediation funding 25 to 1

Bad Actors Already Have the Keys

93%

multi-turn jailbreak success rate on major open weight models (Cisco)

**Zero-day,
AI-built**

a threat actor's AI-developed exploit was caught only because defenders found the bug first

**6 to 12
months**

until many companies have Mythos-class models, possibly released without safeguards (Anthropic's own warning)

The clock is not the safety roadmap of the most careful lab. It is the release schedule of the least careful one.

Machine Speed vs Calendar Speed

2,500 organizations.
106 countries.
Under 1 hour.

One autonomous attack chain run by an LLM behind a custom MCP server.
Initial access, credential dumping, exfiltration: all autonomous.
The only human step was checking the results.

Attackers: minutes

Defenders: change windows, compliance reviews, weeks

Three Moves. This Quarter.

1

Measure patch deployment in hours, not weeks. CISA's three day window will soon look generous.

2

Run frontier models against your own code first. The scan is happening either way. The only variable is who reads the report first.

3

Fund the open source dependencies your stack lives on. The unpaid maintainer triaging your RCE is your most leveraged security spend.

Follow Imiel for more AI signals

Research and analysis on AI agents, regulation, security, and developer tools.
Updated weekly.

imiel.dev