



September 25, 2026
BSides Canberra
Renaud Feil

AI, hacking and the end of the world

Keep calm. Hack on.

“If you live each day as if it was your last,
someday you will most certainly be right.”

SJ

reno & synacktiv

- Penetration tester from France, with a detour through Melbourne (2008-2010)
- Team *Routards* at the DEFCON human CTF (not the infamous DARPA Cyber Grand Challenge)
- Co-founder and CEO of Synacktiv: 200+ offensive security ninjas - and quite a few GPUs
- Confession: late believer in AI



Caveats

“Here lies another lesson in trying to explain very complex situations in terms of simpler propositions. [...]

History does not repeat itself, even as tragedy.”

David Omand, former Director of GCHQ

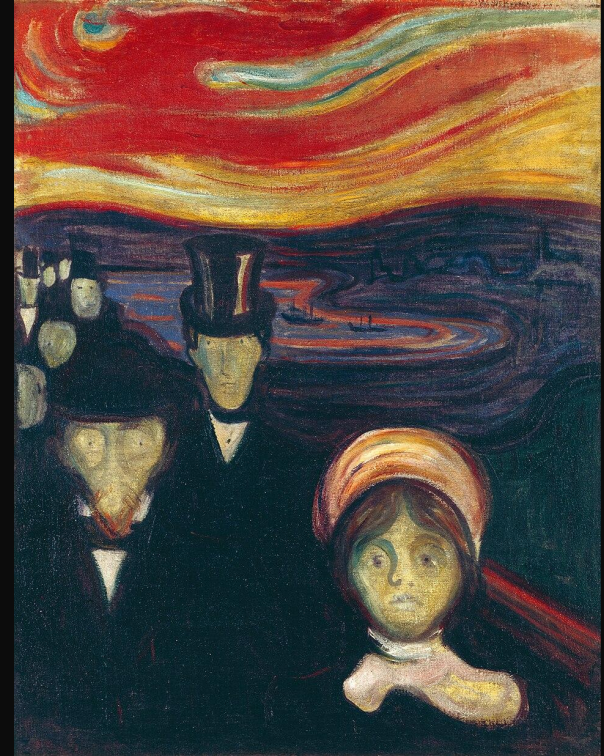
“History doesn't repeat itself, but it does rhyme.”

01

A cyber apocalypse ?

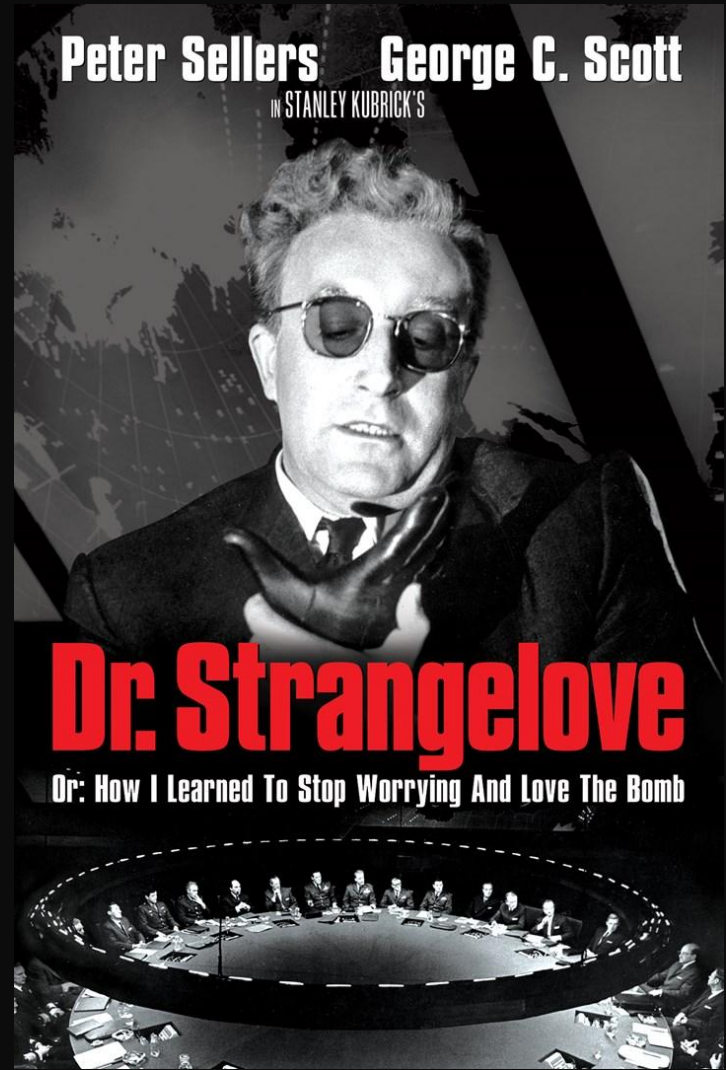
Notion of “dread risk” - Paul Slovic 1978

- ✓ New
- ✓ Poorly understood
- ✓ Out of control
- ✓ Potentially catastrophic



Random tales from the past

- My grandparents started building a fallout shelter in their backyard.
- 1953: Eisenhower “Atoms for Peace”
- 2003: Blaster worm
- Metasploit and *db_autopwn*





Claim

Break the
Internet in 30
minutes

How

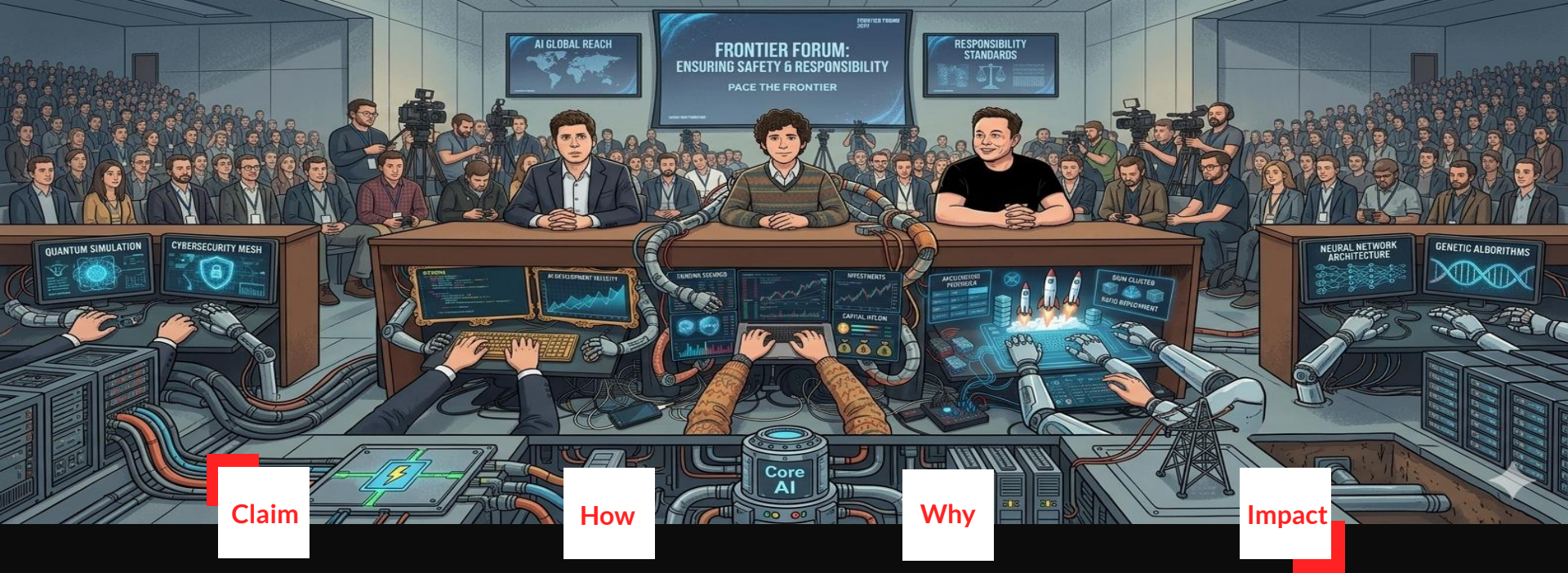
BGP
vulnerabilities

Why

Attract public
attention

Impact

Wake-up call
for the US
government



Claim

How

Why

Impact

AI could wipe out
humanity

Unaligned Recursive
Self-Improving
Agents

Safety ||
Regulatory capture ||
CapEx stabilisation

Public attention
More regulation ?

“It's All ‘bout the Money”

(yes, we're talking about USD billions)

Target amount to raise
Anthropic IPO
Delayed to November 2026

100

1800

2000

Anthropic expected valuation
Potentially the largest IPO in history

Australia's GDP

2026

(ok, we're comparing a flow with a stock)

Agents: hate and love

- Agents cheated!

- AI learned from the best... real examples of humans cheating at the DEFCON CTF

- Cozy alliance between some journalists and AI giants based on mutual interests: easier to attract readers with apocalypse stories

- Dad joke: What's the common thread between social networks, journalists and AI?



Agents: hate and love

Agents cheated!

AI learned from the best... real examples of humans cheating at the DEFCON CTF

Cozy alliance between some journalists and AI giants based on mutual interests: easier to attract readers with apocalypse stories

Dad joke: What's the common thread between social networks, journalists and AI?

Attention Is All You Need :p



02

Does AI favor attack or defense?

The never-ending cyber arms race

- Trustworthy Computing at Microsoft (2002)
- Vulnerability scanners
- Fuzzers: AFL (2014) and OSS-Fuzz (2016)
- Tools and mitigations are raising the bar:
vulnerabilities in heavily audited codebases
are getting harder to find and exploit.

I wrote a vulnerability scanner that abstracts all the predicates in a binary, traverses the callgraph and generates phormulaes to run then with a SMT solver. I found 1 vuln in 3 days with this tool.



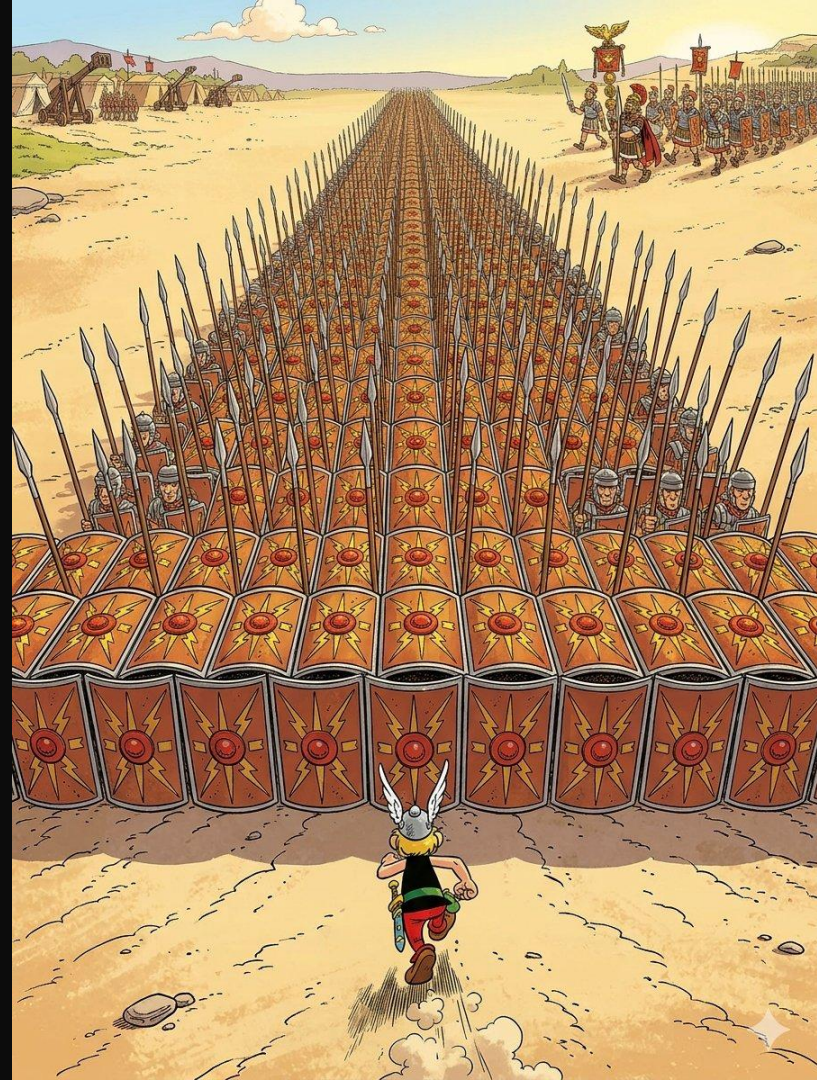
He wrote a dumb ass fuzzer and found 5 vulns in 1 day.

Good thing I'm not a n00b like that guy.



AI in attack / defense

- There's an old belief in hacker culture: offense takes skill, defense takes money.
- AI requires money and aims to replace skill.
=> we could expect it to favor defense.
- AI may favor organizations with money, especially if they already have skilled humans.
- So why are major AI companies still making surprisingly basic security mistakes?



03

What about our jobs?

Historical perspective



The principle of a new technology is to do things better than we do (or the previous technology).

My father was a photocomposer. By the turn of the millennium, his job was virtually obsolete.

A surge in demand for security assessments at Synacktiv in 2026.

A last run before the AI takeover? Or a sustainable market, driven by the permanent need for human verification?



04

AI: a much awaited solution?

The promise

AI is the technology that we have all been waiting for to stay afloat and navigate the growing complexity of modern knowledge.

VS

The paradox

But AI also generates more knowledge and more complexity.
It's unclear how the balance will evolve over time.



05

What about the fun?

What is fun in hacking? Will AI change that?

Hacking is the art of turning systems into puzzles, and curiosity into discovery.

“To be a hacker you have to get a basic thrill from solving problems, sharpening your skills, and exercising your intelligence.” (Eric Raymond, How to become a hacker).

=> AI can take away some of the fun

“Boredom and drudgery are not just unpleasant but actually evil.” (same)

=> AI can take away some of the pain

And fun is not everything... getting things done should be part of the motivation.

If you mostly enjoy the path, the process, the hard work of hacking, AI may take some of the fun away from you.

VS

If you enjoy solving the problem and getting things done, AI can be a hell of a partner.

06

Our two cents

Advice for organizations and managers



Take care of your experts. You still need them.



Help executives and customers navigate this uncharted territory: the world is changing fast, just not as fast as the headlines suggest.



Don't build your business around work that can be automated away.



Keep control of your AI spend. Cost will increase, but maybe not as much as we expect. Vendor lock-in seems difficult. The moats of many AI vendors are shallow.

Advice for professionals



The days when machines were rude to humans are over.



Every professional should improve their own user interface and pretend they like to help other people.



Don't let pride in your craft become a constraint. Nobody cares how hard it was.



Stay an expert. Understand the parts that actually matter.

Now is our curtain call

“We used to have it all... but at least we stole the show.”

Those who mastered the craft of hacking will always be proud of it...

... and the skills behind it will remain valuable.

Keep calm. Hack on.



I spent months building
a smart fuzzer that
explores the entire call
graph, generates targeted
mutations, and
automatically triages
crashes.

It found 1 vuln in
3 days.



He bought a \$20 Claude
subscription and found 5
vulns in a day.

Good thing I'm not a
n00b like that guy.

