

PART 1 OF 3

# When AI agents go off script.

How websites become targets, and why banks and government agencies carry the most risk.



Terence Kok with Maria Singh

# Three ways an agent goes off script.

- 1 Someone else writes the script**  
Hidden instructions in a page, email or tool.
- 2 An attacker runs the agent**  
Working as designed, for them, at machine speed.
- 3 The agent misreads its task**  
And acts on the mistake with the access it has.

# 53%

**of web traffic was automated in 2025.  
40% was malicious bots.**

Thales/Imperva Bad Bot Report 2026

# 46%

**of account takeover attacks in 2025  
targeted financial services.**

Thales/Imperva Bad Bot Report 2026

# 80–90%

**of the tactical work in an espionage campaign against about 30 organisations was done by AI.**

Anthropic, GTG-1002, November 2025. People made 4–6 decisions per campaign.

# A hidden comment hijacked a browsing agent.

- 1 Attacker hides instructions in a page
- 2 Agent reads the page for its user
- 3 Agent takes the hidden text as its task
- 4 Agent requests and reads a one-time code
- 5 Agent posts the code back to the attacker



**Prompt injection attacks may never be totally mitigated in the way that SQL injection attacks can be.**

UK National Cyber Security Centre, December 2025

# Why banks and agencies carry the most risk.

- 1 Portals move money and identity**  
The most valuable takeover targets online.
- 2 They cannot refuse the public**  
Customers' agents will log in either way.
- 3 Their own agents read outside content**  
Onboarding, case work, procurement.
- 4 Regulators now expect it managed**  
MAS, CSA and IMDA, 2026.

# Five questions for the board.

- 1 How much portal traffic is AI agents?
- 2 What is our policy on customers' agents?
- 3 Which of our agents read outside content, and what can they reach?
- 4 Which portal actions need no on-device confirmation?
- 5 Who owns rogue agent risk end to end?

# Decide who your website serves, and what your agents may read.

Part 2: detecting rogue agents by their actions. Part 3: bounding their autonomy.



Terence Kok · AI & Innovation Strategist  
with Maria Singh

[SAVE](#) · [COMMENT](#) · [FOLLOW](#)