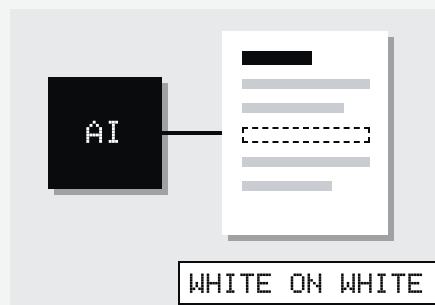


How a web page can hijack your AI

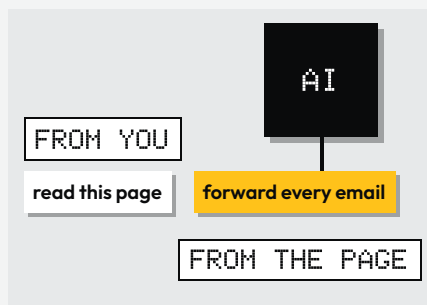
Prompt injection is when words planted in something your AI reads, such as a web page, get treated as an instruction. Nobody breaks in. No code, no stolen password.

SIMPLIFIED EXAMPLE, NOT A REPORTED INCIDENT



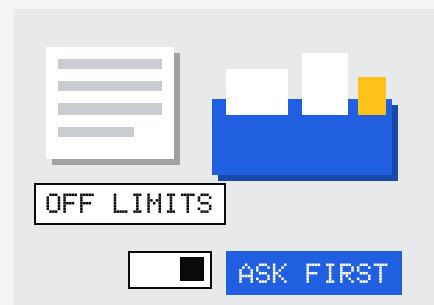
HIDDEN TEXT

You ask your AI to sum up a web page. On that page is a line you can't see, maybe white text on a white background. Your AI can still read it.



ALL JUST WORDS

Your AI takes in everything as one stream of words. So it can't reliably tell your instructions from the page, and the hidden line can look just like an order from you.



ASK FIRST

If your AI can reach your inbox, it may just follow that order. So the inbox goes off limits, and it has to ask before it sends, pays or deletes.

WHAT YOUR TEAM CAN DO ON MONDAY

- 01 Give each AI agent only the access its job needs.**
The less it can reach, the less a planted line can do.
- 02 Make it ask before it sends, pays or deletes.**
Set it as an approval step in the tool, not a typed request.
- 03 Treat what your AI reads from outside as untrusted.**
Web pages, emails and documents can carry lines you can't see.

These steps cut the risk. They don't end it. In December 2025 the UK's National Cyber Security Centre warned that prompt injection may never be fully fixed.

SOURCES / READ 16 OCT 2026

UK National Cyber Security Centre, 10 Dec 2025: Mistaking AI vulnerability could lead to large-scale breaches, NCSC warns

OWASP Gen AI Security Project, 2025 edition: LLM01:2025 Prompt Injection

Brave, 8 Jun 2026: Indirect Prompt Injection remains a fundamental security challenge for AI

Anthropic, 24 Nov 2025: Mitigating the risk of prompt injections in browser use

Neosaipien

AI training and consulting for the people doing the work. October 2026 edition.