

When the page gives the orders: AI agents, web pages and company data

Zspez · October 2026

ABSTRACT

AI browsers and agents now read web pages and act on them. Researchers have shown that a page can steer an agent with hidden text, a crafted link or faint text in an image. Company data is also leaking into public AI tools, often through personal accounts. Banks and large companies have started setting limits. Security agencies say prompt injection may never be fully mitigated. They advise limits, logging and human approval for risky actions. This paper gathers 23 public findings in five short sections. Each cites its source and date. We close with what this means for companies and how Zbrowse will help.

1. Pages can steer AI agents

An AI agent reads a page to act on it. That makes the page an input, and inputs can carry orders.

E1 Hidden text in a forum post made one AI browser read the user's email address and a one-time login code. It then posted both back to the attacker. *(Patched, then reported incomplete.)* [1]

Source: Brave, Aug 20, 2025

E2 Faint text inside a screenshot was enough to give an AI browser new orders. The researchers found the same kind of flaw in several AI browsers. *(Disclosed to the vendors before publication.)* [2]

Source: Brave, Oct 21, 2025

E3 One crafted link made an AI browser send the user's email and calendar data to an attacker's server. *(Disputed by the vendor.)* [3]

Source: LayerX, Oct 4, 2025

E4 The maker of one popular AI browser wrote that prompt injection "is unlikely to ever be fully 'solved.'" In its own demo, a planted email made the agent send a resignation letter to the user's CEO. *(Mitigated in a later update.)* [4], [5]

Source: OpenAI, Dec 2025; TechCrunch, Dec 22, 2025

E5 Gartner told security chiefs to block AI browsers for now. It wrote that "CISOs must block them all until enterprise-ready AI browsers are released in GA to minimize risk exposure." [6], [7]
Source: Gartner, Dec 1, 2025; The Register, Dec 8, 2025

An agent that can click, type and submit needs a person to approve what it does.

2. Company data leaks into public AI

The second risk is quieter. Staff paste work into public AI tools to get help.

E6 Samsung engineers pasted source code and meeting notes into a public AI chatbot for work. [8]
Source: The Korea Herald, May 3, 2023

E7 Samsung then banned generative AI tools on company computers, tablets and phones. [8]
Source: The Korea Herald, May 3, 2023

E8 A bug in one popular AI chatbot showed some users the titles of other users' chats. It also exposed some payment details for 1.2% of paid subscribers. Full card numbers were not exposed. (*Patched.*) [9]
Source: OpenAI, Mar 24, 2023

E9 27.4% of the data employees put into AI tools is sensitive, up from 10.7% a year earlier. (*Vendor data.*) [10]
Source: Cyberhaven, May 2024

Once data leaves the company, the company no longer controls it.

3. Banks are setting agent rules

Banks hold money and identity. They were early to set limits on outside AI.

E10 One of the largest U.S. banks is blocking outside AI agents from directly accessing customer accounts. It still uses AI internally. [11]
Source: CNBC, Oct 13, 2026

E11 The bank's CEO said: "We're not going to allow just any outside agent to come inside the company because that's just too risky." He said customers using outside agents "should understand there's a liability transfer." [12]

Source: CIO Dive, Oct 15, 2026

E12 That bank's terms now set rules for AI agents. No agent may use its systems "unless, at all times, it identifies itself as an AI agent." Agents "must clearly and deterministically label all agentic traffic." [13]

Source: Bank terms of use, accessed Oct 17, 2026

E13 In 2023, several big U.S. banks restricted staff use of public AI chatbots. [8], [14]

Source: The Korea Herald, May 3, 2023; CNN Business, Feb 22, 2023

E14 Other large companies did the same. Two South Korean firms barred AI services internally. A large U.S. retailer warned staff not to share confidential data with a public AI chatbot. [8]

Source: The Korea Herald, May 3, 2023

These are limits on outside agents and public tools. They are not bans on AI.

4. Security experts advise limits and logging

Experts do not promise a fix. They advise designing around the risk.

E15 The UK's National Cyber Security Centre says prompt injection attacks "may never be totally mitigated." It advises least privilege, guardrails built in plain code, and logging of what the model does. [15]

Source: UK National Cyber Security Centre, Dec 8, 2025

E16 OWASP ranks prompt injection the top risk for AI apps. Its advice includes: "Require human approval for high-risk actions." [16]

Source: OWASP GenAI Security Project, 2025 ed.

E17 The U.S. National Institute of Standards and Technology says generative AI "is vulnerable to attacks like prompt injection." It lists data theft among the risks. [17]

Source: NIST, Jul 2024

E18 A U.S. financial regulator reminds firms that its existing rules apply to generative AI. That covers supervision, model risk and data privacy, whether the AI is built in-house or bought. [18]
Source: FINRA, Jun 2024

E19 U.S. health rules allow patient data in a cloud service only with a signed agreement, a risk analysis and security safeguards. (*General cloud guidance, not specific to AI.*) [19]
Source: U.S. Department of Health and Human Services, accessed Oct 17, 2026

5. Employees use AI anyway

Bans alone do not stop use. They can push it out of sight.

E20 78% of AI users bring their own AI tools to work. At small and medium businesses, it is 80%. (*Vendor survey of 31,000 workers in 31 markets.*) [20]
Source: Microsoft and LinkedIn, May 2024

E21 52% of people who use AI at work are reluctant to admit using it for their most important tasks. (*Vendor survey.*) [20]
Source: Microsoft and LinkedIn, May 2024

E22 73.8% of workplace use of one popular AI chatbot runs through personal accounts. (*Vendor data.*) [10]
Source: Cyberhaven, May 2024

E23 In a Samsung staff survey, 65% said security risks can arise from using a public AI chatbot for work. Use was still common enough to prompt a ban. (*Company survey.*) [8]
Source: The Korea Herald, May 3, 2023

Work done in a personal account sits outside company controls and logs.

6. What this means for companies

No browser or agent can promise to stop prompt injection today. Experts say it may never be fully solved. [4], [15] Companies can still limit the damage.

- Keep AI on systems the company controls. [8], [19]
- Make agents ask a person before risky actions. [16]

- Limit what agents can reach, and log what they do. [15]
- Expect sites to set rules for agents, and follow them. [13]
- Give staff a safe tool, so they don't reach for personal accounts. [10], [20]

7. How Zbrowse will help

Zbrowse will be a free, private work browser for company teams. It is in planning and not available yet.

- **AI will run on your company's model or this computer.** Work data will not need to go to a public AI tool.
- **The agent will treat page content as untrusted.** Its actions will stay visible.
- **The agent will ask before every click, form or submit.** A person will approve each step.
- **It will never fill passwords by default.** Passwords will stay yours.
- **Admins will set the rules, and every step will be logged.**

Zbrowse will not make an agent immune to a hostile page. It will keep a person in charge of every action, as the experts advise.

References

1. Brave. "Agentic Browser Security: Indirect Prompt Injection in Perplexity Comet." Aug 20, 2025. <https://brave.com/blog/comet-prompt-injection/>
2. Brave. "Unseeable prompt injections in screenshots." Oct 21, 2025. <https://brave.com/blog/unseeable-prompt-injections/>
3. LayerX (now part of Akamai). "Is Perplexity Comet Safe? LayerX Finds a Prompt Injection Attack Vector." Oct 4, 2025. <https://layerxsecurity.com/blog/cometjacking-how-one-click-can-turn-perplexitys-comet-ai-browser-against-you/>
4. OpenAI. "Continuously hardening ChatGPT Atlas against prompt injection." Dec 2025. <https://openai.com/index/hardening-atlas-against-prompt-injection/>
5. TechCrunch. Report on AI browsers and prompt injection. Dec 22, 2025. <https://techcrunch.com/2025/12/22/openai-says-ai-browsers-may-always-be-vulnerable-to-prompt-injection-attacks/>
6. Gartner. "Cybersecurity Must Block AI Browsers for Now." Research note, Dec 1, 2025. <https://www.gartner.com/en/documents/7211030>

7. The Register. Report on Gartner's AI browser advice. Dec 8, 2025.
<https://www.theregister.com/software/2025/12/08/block-all-ai-browsers-for-the-foreseeable-future-gartner/2793276>
8. The Korea Herald. "Samsung bans AI chatbots after leak." May 3, 2023.
<https://www.koreaherald.com/article/3118116>
9. OpenAI. "March 20 ChatGPT outage: Here's what happened." Mar 24, 2023.
<https://openai.com/index/march-20-chatgpt-outage/>
10. Cyberhaven. "Q2 2024 AI Adoption and Risk Report." May 2024.
<https://info.cyberhaven.com/hubfs/Content%20PDF/Cyberhaven%20Q2%202024%20AI%20Adoption%20and%20Risk%20Report%20052024.pdf>
11. CNBC. Report on a large U.S. bank and outside AI agent account access. Oct 13, 2026.
<https://www.cnbc.com/2026/10/13/jpmorgan-ai-agents-account-access.html>
12. CIO Dive. "Top banks reckon with AI risks, but spot benefits." Oct 15, 2026.
<https://www.ciodive.com/news/top-banks-ai-risks-benefits/833052/>
13. Bank terms of use, section "AI Agent or Agentic AI Notices." Accessed Oct 17, 2026.
<https://www.chase.com/digital/resources/terms-of-use>
14. CNN Business. Report on a large bank restricting staff chatbot use. Feb 22, 2023.
<https://www.cnn.com/2023/02/22/tech/jpmorgan-chatgpt-employees>
15. UK National Cyber Security Centre. "Prompt injection is not SQL injection (it may be worse)." Dec 8, 2025. <https://www.ncsc.gov.uk/blog-post/prompt-injection-is-not-sql-injection>
16. OWASP GenAI Security Project. "LLM01:2025 Prompt Injection." 2025 ed.
<https://genai.owasp.org/llmrisk/llm01-prompt-injection/>
17. NIST. "AI RMF: Generative AI Profile" (NIST AI 600-1). Jul 2024.
<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf>
18. FINRA. Regulatory Notice 24-09. Jun 2024.
<https://www.finra.org/sites/default/files/2024-06/regulatory-notice-24-09.pdf>
19. U.S. Department of Health and Human Services. HIPAA FAQ 2075, on using a cloud service to store or process ePHI. Accessed Oct 17, 2026. <https://www.hhs.gov/hipaa/for-professionals/faq/2075/may-a-hipaa-covered-entity-or-business-associate-use-cloud-service-to-store-or-process-ephi/index.html>
20. Microsoft and LinkedIn. "AI at Work Is Here. Now Comes the Hard Part." 2024 Work Trend Index, May 2024. <https://www.microsoft.com/en-us/worklab/work-trend-index/ai-at-work-is-here-now-comes-the-hard-part>