Mistaking AI vulnerability could lead to large-scale breaches, NCSC warns

NCSC raises alert on "dangerous" misunderstanding of emergent class of vulnerability in generative artificial intelligence (AI) applications.

- Latest blog warns against comparing prompt injection attack vector to SQL (Structured Query Language) injection due to crucial differences.
- Developer community urged to focus on secure design over silver-bullet solutions to help mitigate risks from "inherently confusable" AI-driven systems.

The National Cyber Security Centre (NCSC) – a part of GCHQ – has shared critical insights cautioning cyber security professionals against comparing prompt injection and more classical application vulnerabilities classed as SQL injection.

A new blog advises that, contrary to first impressions, prompt injection attacks against generative artificial intelligence applications may never be totally mitigated in the way SQL injection attacks can be.

Unlike SQL mitigation techniques, which hinge on enforcing a clear separation between data and instructions, prompt injection exploits the inability of large language models (LLMs) to distinguish between the two.

Without action addressing this misconception, the NCSC warns, websites risk falling victim to data breaches exceeding those seen from SQL injection attacks in the 2010s, impacting UK businesses and citizens into the next decade.

Backing proactive adoption of cyber risk management standards, the NCSC challenges claims that prompt injections can be 'stopped'.

Instead, it suggests efforts should turn to reducing the risk and impact of prompt injection and driving up resilience across AI supply chains.

As AI technologies become embedded in more UK business operations, the NCSC calls on AI system designers, builders and operators to take control of

manageable variables, acknowledging that LLM systems are “inherently confusable” and their risks managed in different ways.

Read the blog in full

PUBLISHED

10 December 2025

WRITTEN FOR

Cyber security professionals

NEWS TYPE

General news