Installing an agent component is not a trust ceremony. It is an authority transfer.

Dr Peter McCann Strain
CTO, DPhil/PhD in AI from Oxford University

Swipe ››

— THE INCIDENT PATTERN

An agent dependency is not only code. It can be tool descriptions, skills, configs, manifests and prompts: text the model reads as instruction. The old question was what code you pulled into your process. The new one is what text you pulled into your model's authority.

— THE REFRAME

Admission is authority transfer.

THE OLD QUESTION

What code did you pull into your process?

↓ ↓ ↓

THE QUESTION THAT HOLDS UP

What text did you pull into your model's authority?

— WHAT THE RECORD SHOWS

Roughly a third of skills flagged

Snyk's February 2026 ToxicSkills research scanned 3,984 skills from two public registries, ClawHub and skills.sh. Roughly a third carried a security flaw; one in eight a critical issue. 76 were hand-verified as malicious payloads. The proportions won't hold to the decimal; the contour will.

SOURCE

Snyk's ToxicSkills report; with Invariant Labs' MCP tool-poisoning disclosure, NVD's CVE-2025-54136 for Cursor, and the LiteLLM MCP command-execution advisory.

FAILURE CHAIN

Run the 4P Gate before any component enters an agent environment.

01 **Provenance** and **Permissions**: who made the exact version, and what can it reach?

02 **Posture**: what scan, config diff or hidden-text check ran before approval?

03 **Pull-direction**: can it fetch new instructions or commands after approval?

— THE ARTIFACT

Four gates at the door: Provenance, Permissions, Posture, Pull-direction.

Provenance
Where did it come from?

Permissions
What can it reach?

Posture
What risks are known?

Path
Can it fetch new instructions?

Before any new tool, skill or connector enters an agent environment, run the 4P Gate. Any 'no' is a stop sign.

— DO THIS AFTER THE NEXT INCIDENT

Pick one MCP server, skill, connector or agent config installed in the last thirty days. Answer the 4P Gate in writing. Name which P blocks first, who owns that block, and the date you will re-run it.

FAILURE MODE TO AVOID

Pinning install approval to a signed publisher and a licence check. Provenance covers neither hidden instructions in tool descriptions nor what the component can fetch after approval. Run all four P gates, not the familiar two.

— USE THE FULL POSTMORTEM

The Supply Chain You Cannot See

Read the full essay — the argument, the sources, the figures and a reader-ready working artifact.
Substack petermccannstrain.substack.com · Medium @peter.mccann.strain · LinkedIn peter-strain-dphil-15a607128
New essays twice weekly, 2 June – 21 July 2026.

Next: E11 — The Sentence That Owns the Agent

— THE STACK SO FAR

E10 · Essay 10 of 22 complete · Arc II: Evidence and authority

YOU JUST ADDED

The 4P Gate and Agent Instruction BOM

STACK LAYER LIT UP

Tools / Permissions

YOU CAN NOW ASK

inspect instruction-bearing components before they enter the agent.

NEXT

E11 asks how a single sentence can capture an agent's authority at runtime.

Dr Peter McCann Strain

CTO, DPhil/PhD in AI from Oxford University

I build production AI systems and write about making agentic AI useful, inspectable, governable and safe enough for real work.

Follow on Substack for the full 22-essay series
petermccannstrain.substack.com