

BLAST RADIUS

Containing one agent's mistake.

Delegated authority enforced at governed action boundaries, limiting what a compromised agent can do.

WHAT WE KEEP HEARING

The concern isn't a bad answer. It's an agent acting outside its remit, or permissions being widened or exploited further down a workflow.

With several agents working together, firms need to verify identity, keep guardrails applied end to end and contain a compromised agent. The exposure is the lethal trifecta: private data, untrusted content and the ability to communicate externally.

WHAT NUGGETS DOES

Nuggets intercepts governed actions and checks delegated authority before execution, at the handoffs and action boundaries brought into scope, rather than only at workflow start.

Within that scope, an agent cannot widen its permissions or inherit ungranted authority by passing work to another agent. Out-of-policy actions are denied or escalated to a human.

Manipulated input may change what an agent attempts, but authority checks sit outside the agent and its runtime. Governed actions must still pass the firm's policy, and permissions can be revoked.

PILOT SCOPE

Two agents, one attempting to escalate privileges or pass an unauthorised action through the other. Show the denial, the containment and the receipt for both.

OUTCOME

Blast-radius containment.