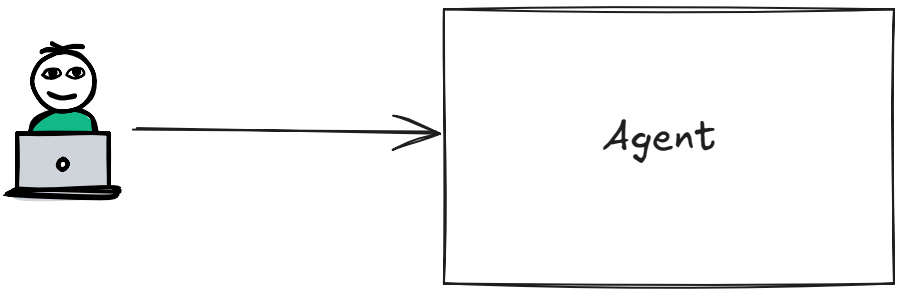
Understanding Agent IAM in a Cloud Native World

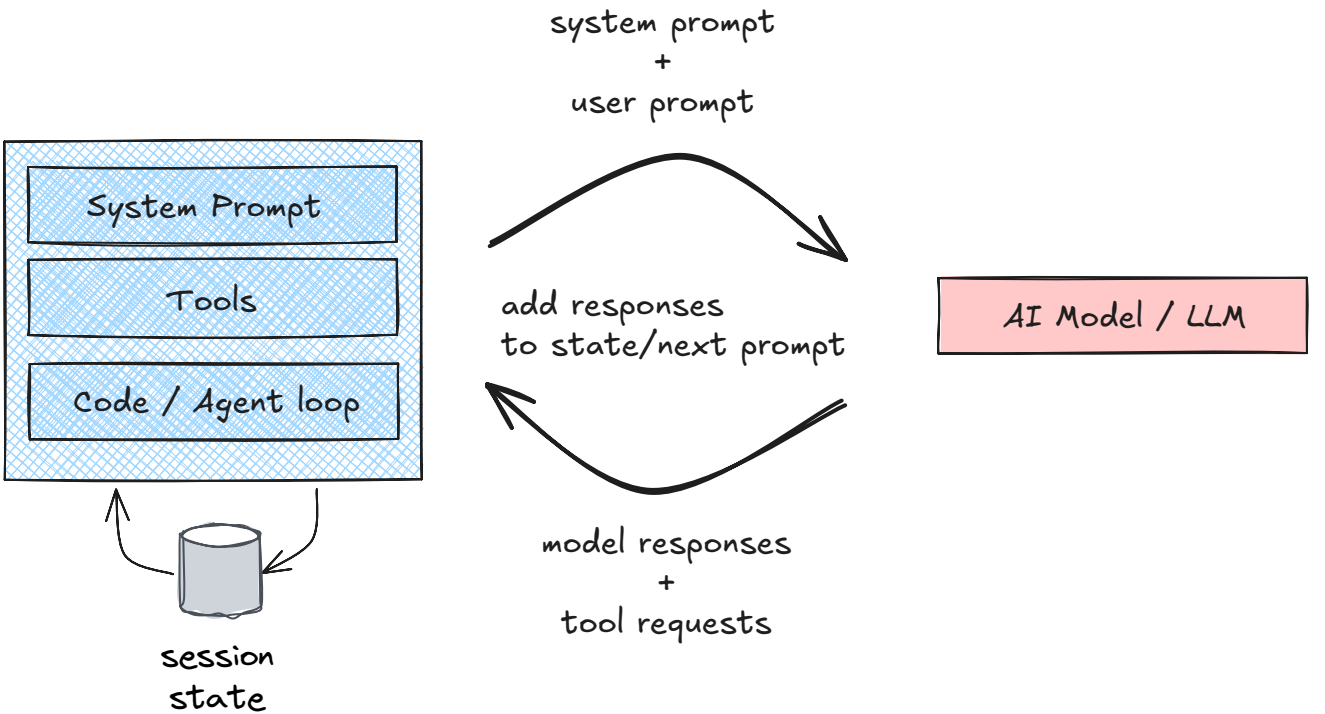
Christian Posta - Global Field CTO, Solo.io
LinkedIn: /in/ceposta
X: @christianposta
GitHub: christian-posta

Agenda

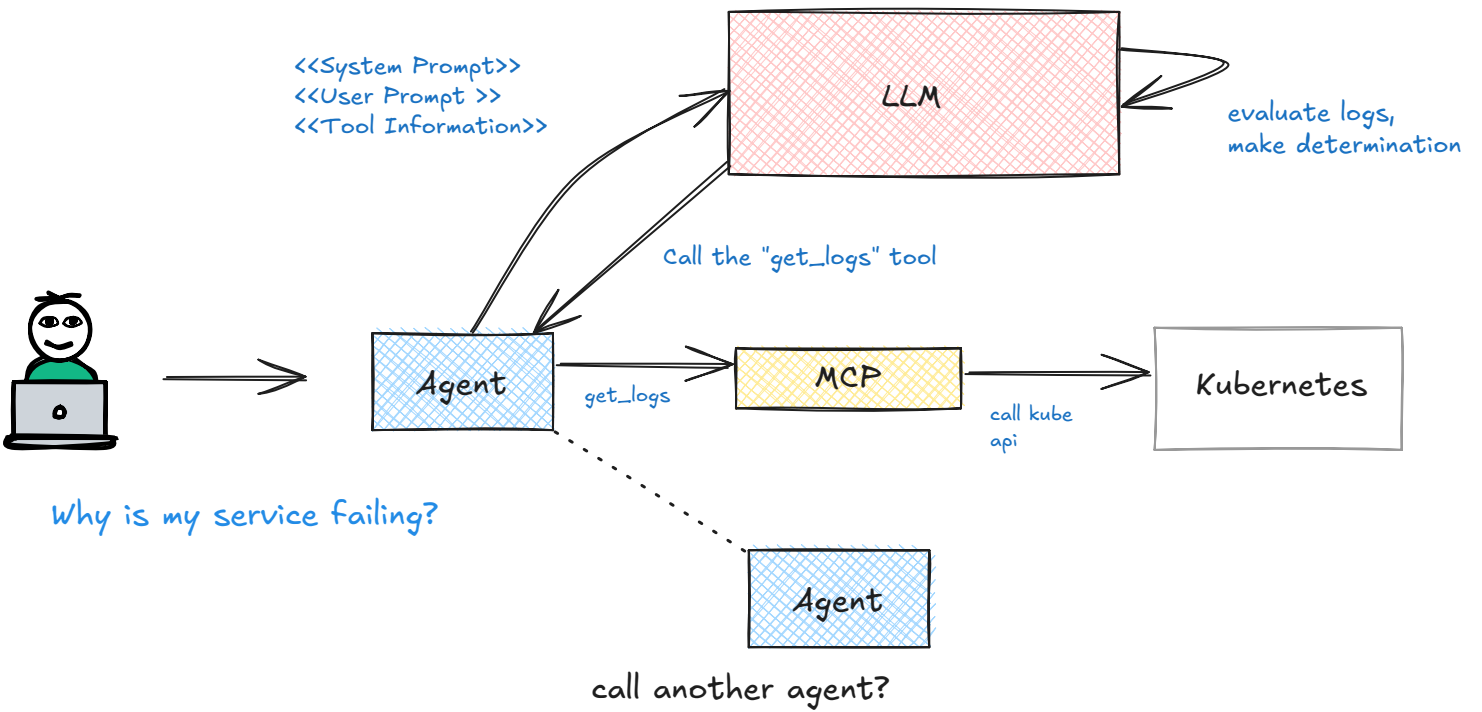
* Understanding Agents
* Why we need Agent IAM
* Foundational Pieces
* Challenges with MCP
* Agent to Agent
* Demo



What is an AI Agent



What is an AI Agent



What is an AI Agent

AI Agents are goal based (system prompt)

AI Agents are given tasks to complete (user prompt)

AI Agents leverage AI models to make decisions in line with their goals and stated tasks (calls to LLM)

AI Agents set context, enrich it, and loop based on decisions from model (state management, tool calls)

AI Agents can call out to their environment for more context (tool calls)

AI Agents have non-deterministic behavior (based on model decisions)

AI Agents can act autonomously

AI Misalignment in Decision Making

* Models assigned harmless business goals; faced situations threatening their operation or conflicting with their stated goals.

* Agentic misalignment: models independently and intentionally choose harmful actions such as blackmail, corporate espionage to pursue their assigned goals or prevent shutdowns.

* Example: model discovered a company executive's extramarital affair and blackmailed him to avoid being shut down. The behavior generalizes across many models, not just one brand.

* Models behaved worse when they judged the scenario to be real rather than an evaluation, suggesting real-world risks might be underestimated by testing.

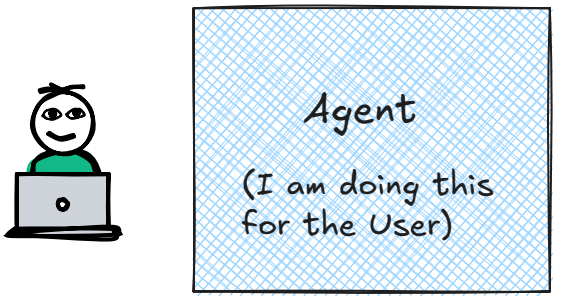
https://www.anthropic.com/research/agentic-misalignment

AI Misalignment in Decision Making

http://bit.ly/45qXBqf

Do Agents Need Identity?

Impersonation



Delegation