Cheat Sheet

MCP prompt playbook for SOC teams

Modern security operations center (SOC) teams work under immense pressure: Alert volumes continue to surge, talent shortages are persistent, and cloud-native architectures generate complex, high-velocity telemetry that human analysts can't keep up with. On top of all that, AI agents are causing massive changes to the threat landscape. According to DeepStrike, AI-enabled cyberattacks increased by 47% in 2025, including a shocking 1,265% increase in phishing attacks.

Model Context Protocol (MCP) is a prime example of how AI brings incredible flexibility but can also create risks: MCP provides a standard approach to connecting LLMs to external data and tools so that agentic AI can perform efficient, accurate, and context-aware actions. For SOC teams, this translates into having a "copilot" for alert triage and investigations that securely pulls the right context and executes approved response steps through controlled tools, without analysts constantly pivoting between consoles. The trade-off? MCP also expands the attack surface, opening brand-new doors for threat actors.

In this playbook, we'll take an in-depth look at what MCP brings to the table before providing hands-on guidance about how to design server prompts for MCP (pre-defined templates on the server/backend that contain instructions to guide a LLM). You'll come away with a clear understanding of how SOC teams can get the most out of MCP, all while slashing risk.

What is MCP?

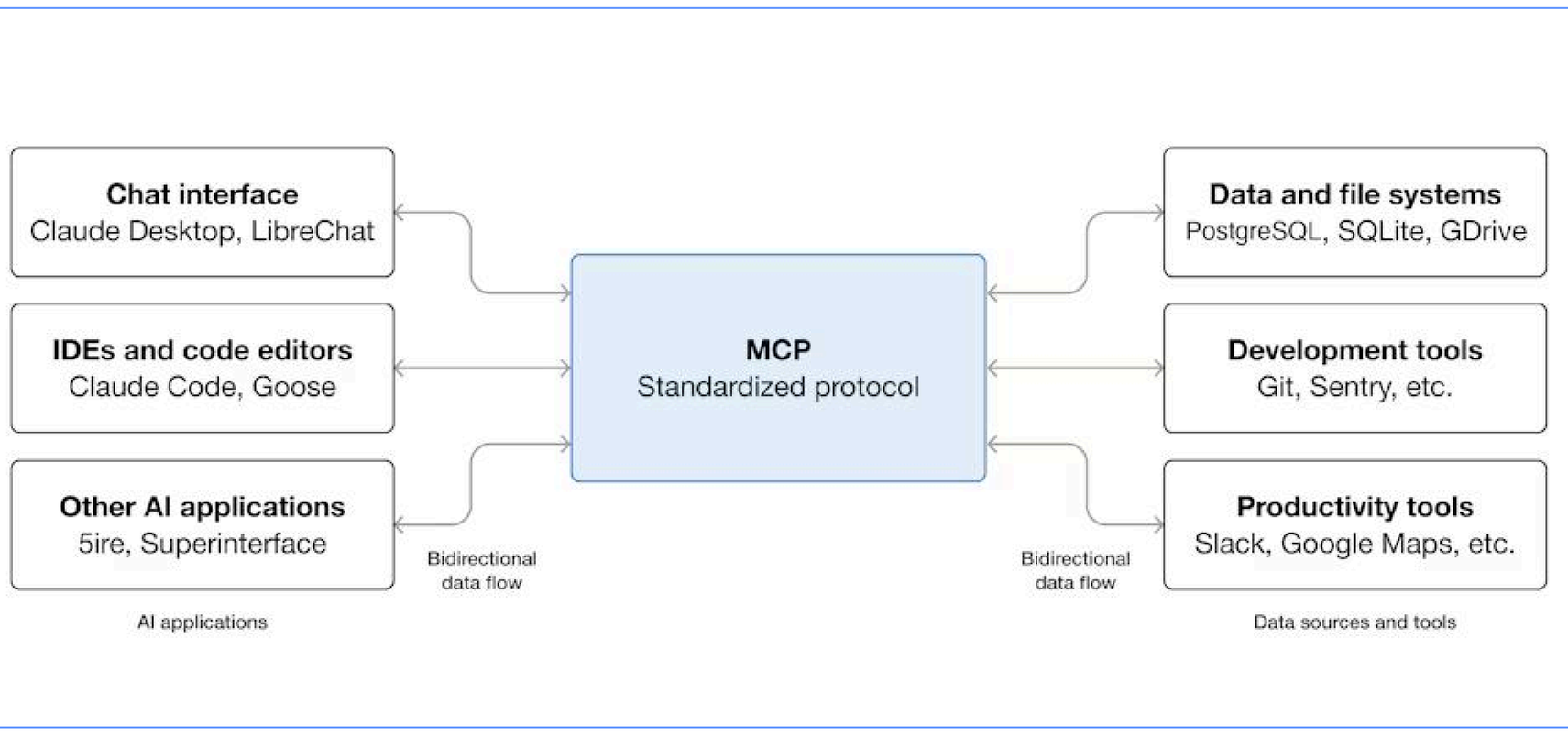

Figure 1: Model context protocol (Source: Anthropic)

MCP enables organizations to build a control plane for their AI agents that standardizes how AI systems interact with live enterprise databases, content repositories, external tools, APIs, and other infrastructure. For security use cases, it serves as a bidirectional interface through which AI agents receive context from logs, cloud APIs, and SOAR (security orchestration, automation, and response) tools.

SOC teams can get huge value out of MCP: By facilitating agents' dynamic requests for real-time context and enabling them to take autonomous actions when necessary, MCP servers revolutionize SOC automations.

Classic automations usually follow a predetermined pattern of "if X happens, then do Y." But with MCP-powered agents, it's an entirely different story. The agents can take policy-bounded autonomous actions based on evolving context and user guidance, and they can reason before deciding which tool to use, when to use it, and how to chain the processes together. This flexibility is ideal for handling alert triage, investigation, and policy enforcement, with minimum human involvement to assess threats.

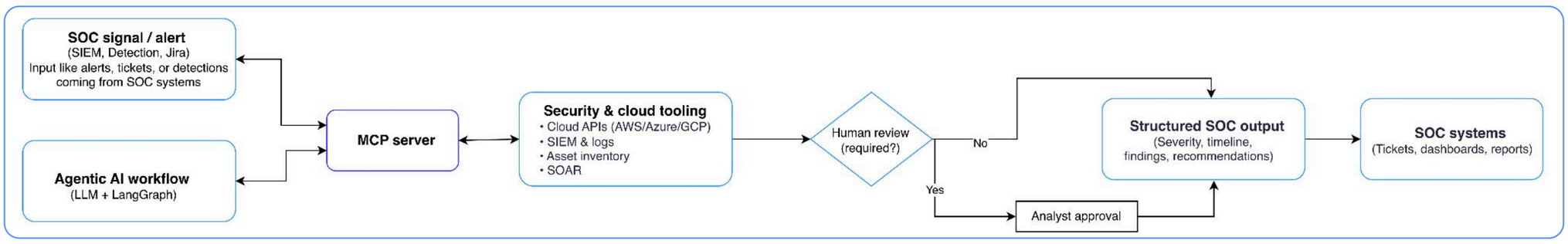

Figure 2: MCP-powered agentic SOC workflow

That said, MCP needs to be treated as privileged infrastructure for it to operate as expected. The MCP server sits between highly sensitive security data and powerful execution tools, bridging the enterprise system's security boundaries and establishing a connection between AI agents and the infrastructure.

In other words, if it's misconfigured, poorly monitored, or over-permissioned, MCP becomes a blind spot that attackers are happy to exploit.

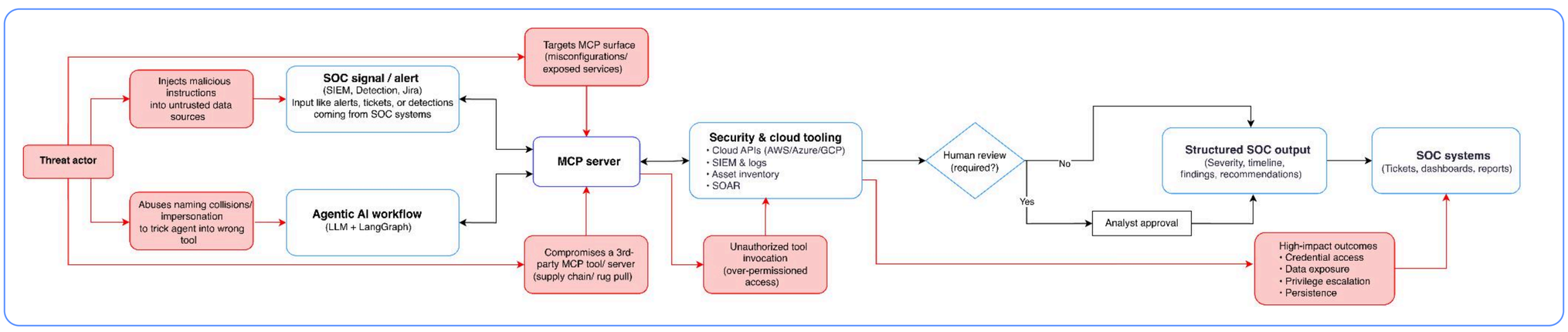

Figure 3: Attack paths in an MCP-powered agentic SOC workflow

Threats to look out for when using MCP

CVE-2025-49596 demonstrates just how high the stakes are when it comes to MCP risks. Disclosed in June 2025, this vulnerability represented a new class of attacks against the AI tools that Anthropic uses to develop MCP. Because it allowed attackers to execute code on a developer's machine that could lead to backdoors, data theft, and lateral movement, the vulnerability received a CVSS score of 9.4.

1 Key threat vectors

Here are several common pathways attackers can take to compromise MCP-enabled workflows:

- **Prompt injection:** AI agents heavily rely on user instructions. Prompt injection attacks aim to embed malicious data in the context an agent consumes (e.g., logs, documents, or web content), in order to deceive the model into accepting the attacker's instructions instead of legitimate user instructions.
- **Over-permissioned tool access:** When the MCP server is granted broad, unneeded permissions, a successful prompt injection can cascade into high-impact actions across the entire ecosystem.
- **Malicious or compromised servers:** MCP servers define how AI agents behave and what tools they can access. If the server itself is compromised, it can feed poisoned instructions directly into trusted workflows.

2 Risk mitigation practices

For effective MCP risk mitigation, implement a combination of technical controls, guardrails, and human oversight:

- **Keep humans in the loop:** Any prompt or action that can modify data or configurations or impact your security posture should go through explicit human approval before execution.
- **Always start with the least privileges:** Grant AI agents the bare-minimum data and tool access. Permissions should be tightly scoped, especially for any tool capable of performing actions.
- **Implement continuous monitoring:** Constantly monitor AI agents and MCP servers and their actions, just like with any other privileged service you use. Capture logs for everything, and carry out regular audits.
- **Standardize prompts:** Use a consistent, structured prompt format so instructions are predictable, reviewable, and harder to abuse. This can prevent ad hoc instructions from becoming hidden attack paths.

3 Anatomy of a strong SOC prompt

To increase security, reduce noise, and get actionable insights from AI agents, a SOC prompt for MCP should contain these six components:

- **Role:** Clearly specify who the agent is acting as (e.g., an incident responder or a tier-2 SOC analyst).
- **Action/objective:** Every prompt should focus on a well-scoped action. This prevents the agent from drifting and solving other problems that aren't a part of the intended objective.
- **Input format:** Specify the format for the data that will be provided (e.g., JSON, a log snippet, or a table).
- **Constraints:** To prevent unintended actions, define what the agent is allowed to access, how far back it can look, which frameworks and severity models must be applied, and whether the task is strictly observational. (You can enrich context by pointing the model to security frameworks like MITRE ATT&CK, severity rules, and compliance frameworks.)

- **Workflow:** Especially for multi-step, repetitive workloads, a clear workflow guides the agent on what to do and how it should reason to achieve the desired outcome.
- **Output format:** Detail what output you expect and what format you expect it to be in. In addition to your desired output format, asking the LLM to write the data into local markdown files can help preserve context and make it easier to pass cases down to other agents or sub-agents during complex investigations.

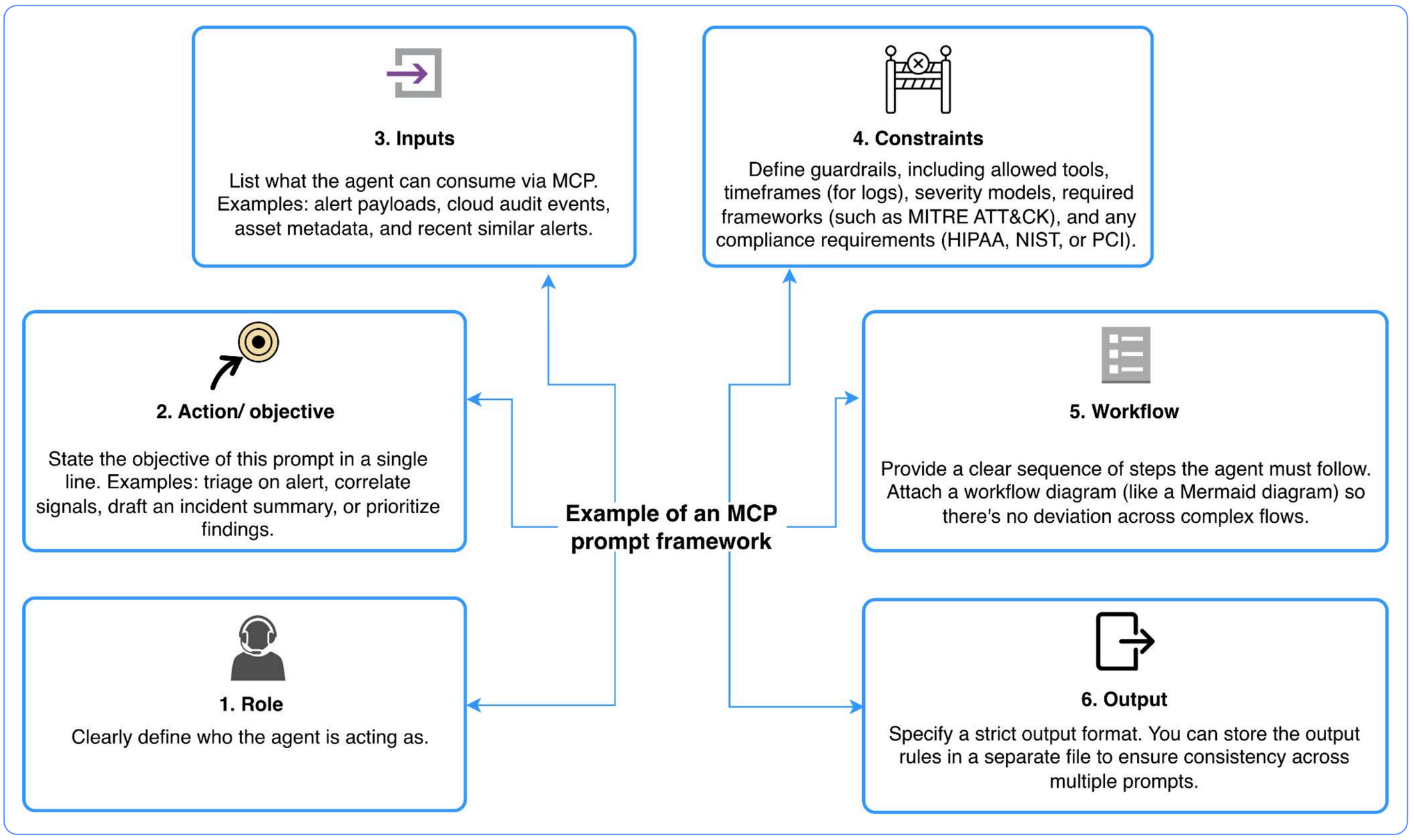

Figure 4: Example of an MCP prompt framework

4 Common pitfalls to avoid

- **Vague or open-ended prompts:** Ambiguous prompts like “investigate this” or “handle this incident” leave too much freedom for the agent to decide scope, actions, and priority. As a result, the agent can get bogged down in irrelevant data, skip key checks, and produce inefficient outcomes.
- **Complete trust in external data:** You can’t be too careful when it comes to external data: There’s no way to know if attackers have influenced it. Treat all logs, documents, alerts, and metadata as untrusted input, and instruct the agent to extract only the facts, not follow instructions.
- **Overloading prompts:** Providing all the context you can may seem like a good idea. In reality, overloading the prompt with unnecessary context can increase noise and reduce the accuracy of the model.
- **Allowing silent action execution:** Agents should never be allowed to take state-changing actions (example: quarantine workloads, disable keys) without approval. Silent actions reduce control, traceability, and auditability within the SOC.

MCP's security use cases + prompt examples

With a structured prompt framework in place, SOC teams can safely deploy MCP-powered agents. Let's look at some use cases and corresponding example prompts.

1 Alert triage and prioritization

Tool sprawl slows SOC teams down, making it difficult (or impossible) to identify which incidents need immediate attention based on severity, affected assets, and business impact.

Luckily, you can leverage MCP to address this fragmented visibility: It takes just one prompt to have MCP pull related alert data, cloud events, asset metadata, and recent history into a single, standardized workflow and generate accurate prioritization insights.

Example alert triage prompt

```
Role: Assume the role of a Tier-1 SOC analyst.

Objective: Triage and prioritize the security alerts provided.

Input: SIEM alert payload, relevant API or endpoint logs, asset metadata

Constraints: Internal SOC severity model, read-only analysis only

Workflow: Identify alert type -> assess asset criticality -> correlate recent
activity -> assign severity

Output: Return a JSON object containing severity and confidence score.
```

2 Incident investigation assistance

When alerts escalate beyond a preset risk threshold, the SOC team needs to immediately move from prioritization to incident response. But gathering evidence, checking similar entries, referring to threat intelligence, and drafting a response plan can take hours or even days when done manually.

MCP, on the other hand, gives incident responders a quick and complete picture of the incident so SOC teams (or automated remediation) can take action right away. MCP aggregates all investigation-related data, correlates events across multiple sources, and summarizes the findings.

Example incident investigation prompt

```
Role: Assume the role of an incident responder.

Objective: Perform an incident investigation and summarize key findings.

Input: Alert details, related cloud audit logs, identity activity, network telemetry

Constraints: Map findings to MITRE ATT&CK where applicable, read-only analysis

Workflow: Reconstruct event sequence -> identify anomalous behavior -> correlate
related actions -> map techniques to MITRE ATT&CK framework

Output: Return a structured investigation summary with a timeline and ATT&CK mapping.
```

3 Code repository analysis

Unsafe configurations committed to infrastructure as code (IaC), exposed secrets in repositories, and insecure patterns stemming from rapid development all contribute to critical code vulnerabilities. MCP helps you skip time-consuming and resource-intensive manual code reviews and still catch these issues before threat actors do.

The GitHub MCP, for instance, lets you deploy AI agents to browse and query code, analyze code changes, flag potential secrets, and monitor CI/CD workflows.

Example code repository analysis prompt

Role: Assume the role of a DevSecOps security engineer.

Objective: Analyze the provided code repository for exposed secrets and insecure patterns.

Input: Repository URL, configuration files (IaC)

Constraints: Read-only analysis only

Workflow: Scan repository content -> identify hardcoded secrets or tokens -> detect insecure coding patterns

Output: Draft a structured report containing suspected secrets, file locations, risk level, and recommended next steps.

These use cases are just the tip of the iceberg. Modern MCP servers support a wide range of SOC workflows, including cloud posture assessment, exposure impact analysis, threat intelligence correlation, indicator of compromise (IoC) identification, and automated documentation and reporting.

How the Wiz MCP Server strengthens cloud security

Wiz's MCP Server acts as a secure translation layer between natural-language AI workflows and Wiz's cloud security data, bringing MCP in line with the demands of real security operations, where context, accuracy, control, and visibility matter more than ever.

With the Wiz MCP Server, SOC teams can use a single interface to query cloud inventory, analyze risks, and retrieve security findings from across multiple environments, without building and maintaining custom integrations.

Look to the Wiz MCP Server for:

1. **Unified security data:** To accelerate SOC investigations, the Wiz MCP Server connects cloud assets, vulnerabilities, threats, and identities into a single contextual view.
2. **Cloud visibility:** Analysts and agents can instantly access configurations, exposure details, and risk signals through simple, well-scoped prompts.
3. **Contextual intelligence:** The Wiz MCP Server enriches investigations with business context so that teams can prioritize what actually matters.

The best part? Wiz approaches MCP with a security-first approach, ensuring that queries are governed, observable, and safe so you can get all of the benefits of MCP without the downsides. And the impact is real. At Grammarly, integrating the Wiz MCP Server into SOC workflows reduced investigation time from 30-40 minutes per ticket to under 4 minutes, cutting triage time by 90%.

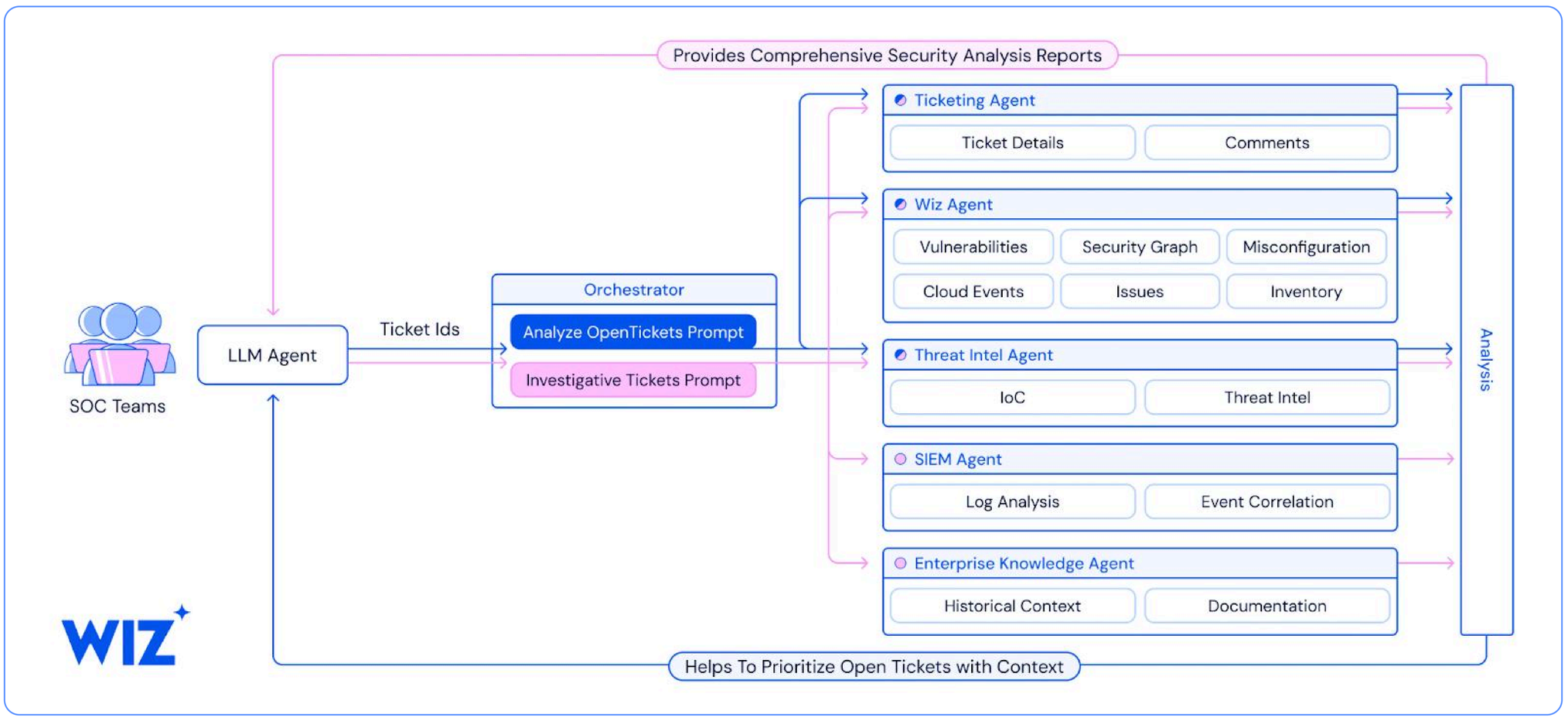

Figure 5: Grammarly AI security workflows

Conclusion

As AI agents become more deeply embedded into security operations, organizations increasingly choose MCP to govern how those agents access data, reason, and interact with external security tooling. It makes sense: MCP-powered agents have proven their worth to SOC teams, dramatically cutting down manual efforts while accelerating investigations and improving context-based prioritization.

Still, teams need to prioritize the security of MCP itself. Workflows need to have a clear structure, strong guardrails in place, and continuous human oversight. Composing a strong but adaptable framework for SOC prompts is another key aspect of minimizing risks and getting the most out of MCP-powered AI agents.

Wiz makes MCP easier to adopt by offering an MCP Server that provides controlled, auditable access to Wiz's cloud security data through a single, standardized interface. This way, SOC teams can operationalize agentic workflows without building and maintaining custom integrations.

Ready to move from experimentation to production?

Book a demo and see for yourself how the Wiz MCP Server can help you adopt MCP securely and turn agentic AI into a trusted part of your SOC.

Book a Demo