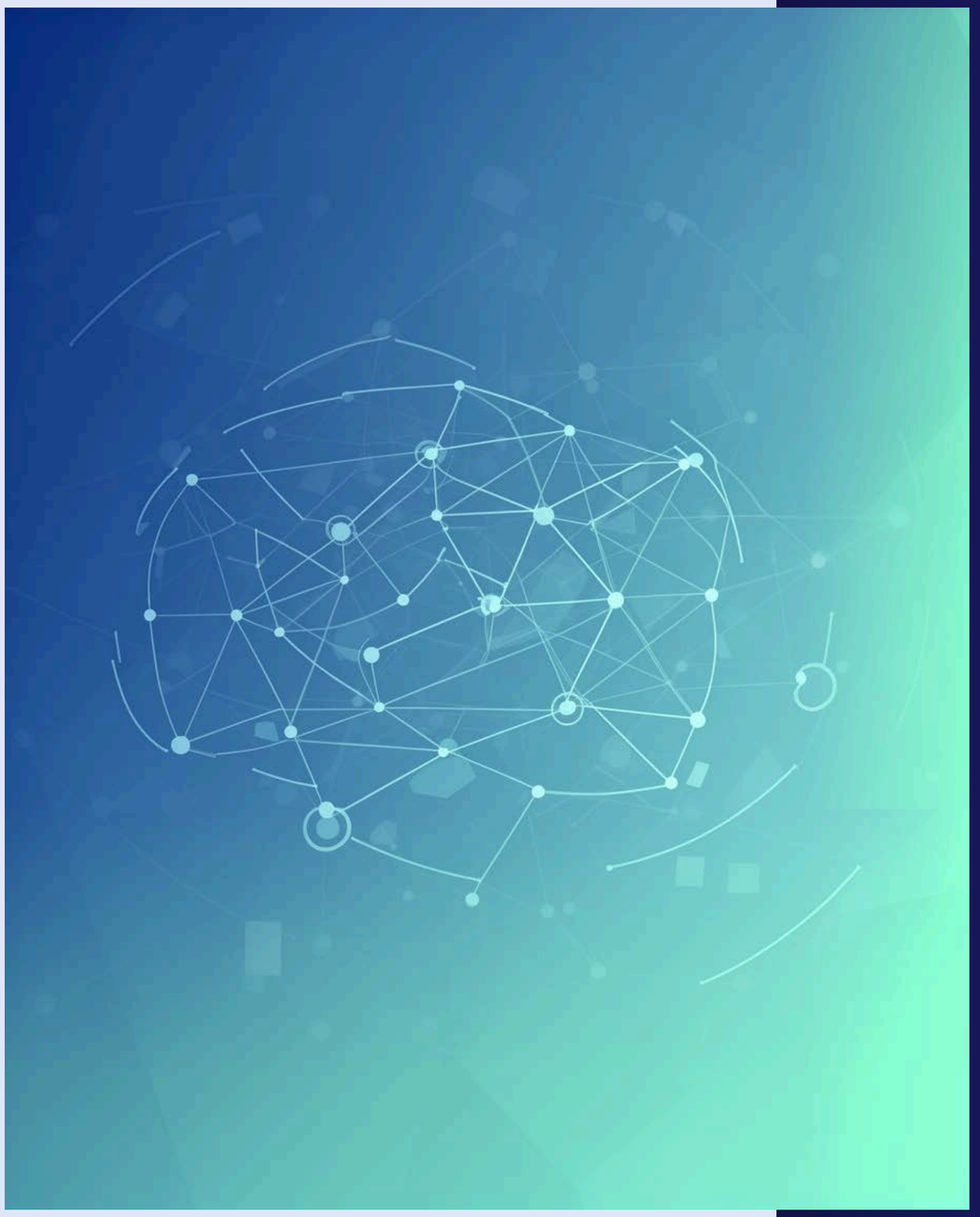


AI-Assisted Incident Response Using LLMs and MCP in Distributed Systems

Makarand Gujarathi
Individual Researcher

SRE 2026

The SRE Challenge

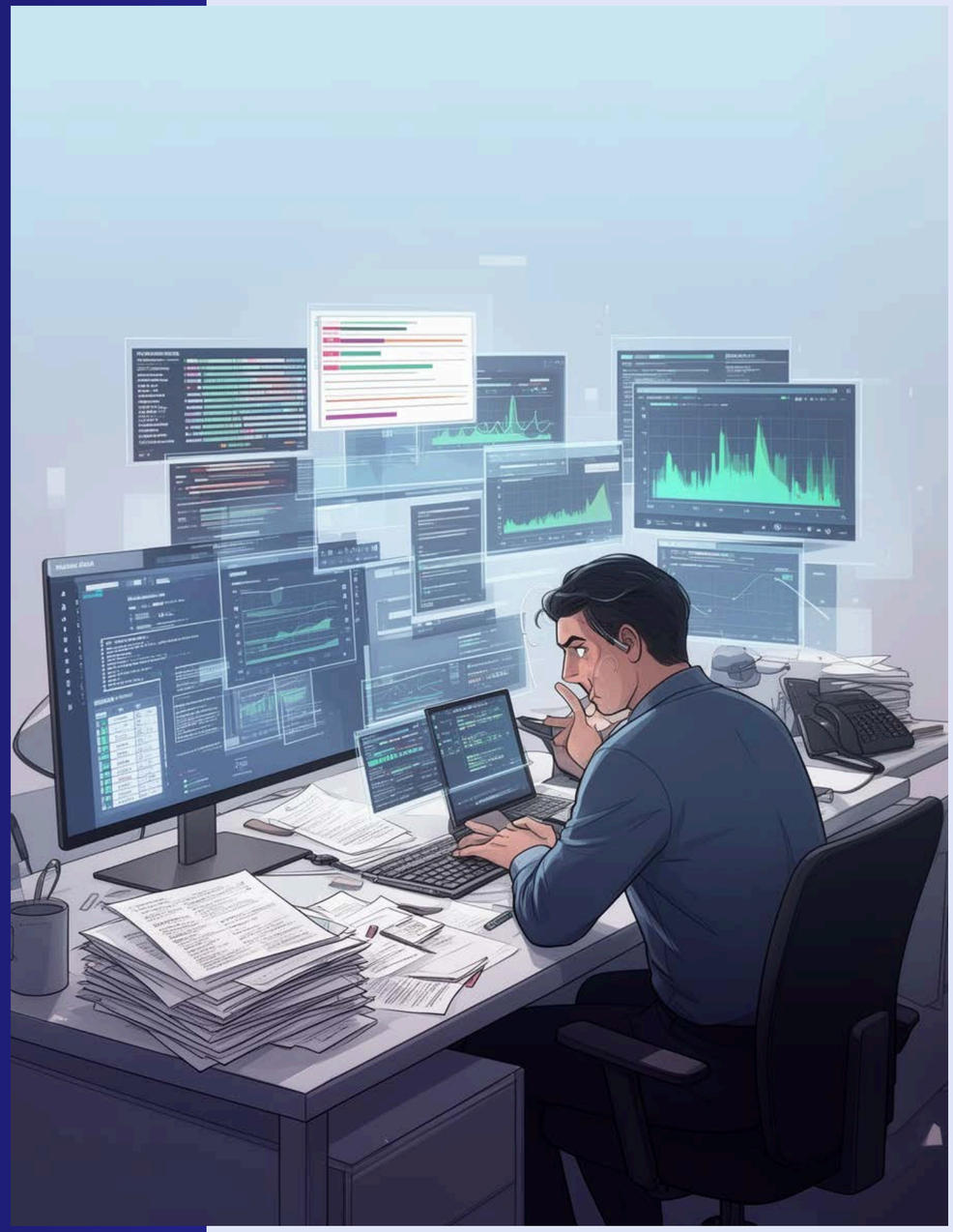
DIAGNOSING INCIDENTS IN MICROSERVICES ENVIRONMENTS

Diagnosing incidents in microservices environments requires understanding interactions between 5–10 backend components, which leads to challenges in telemetry and incident resolution processes.

Incident Investigation Pain Points

Challenges in Current Workflows

- Engineers spend 2–4 hours
- Manual correlation across stores
- Expertise takes months to acquire
- Overwhelming logs and metrics
- Limited visibility slows response



Introducing AI-Assisted Incident Response

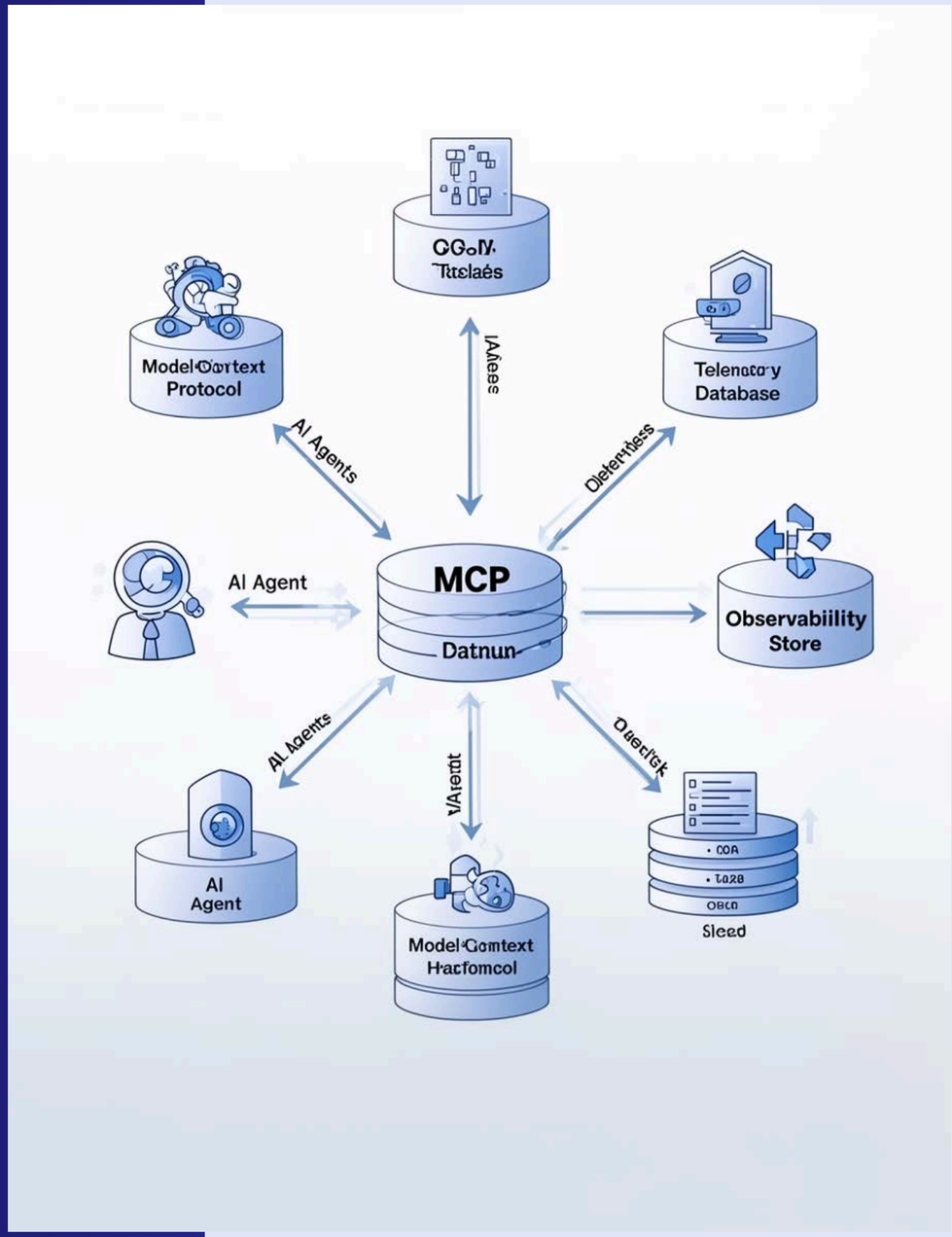
Leveraging LLMs and MCP Technology

- Large Language Models (LLMs) integration
- Model Context Protocol (MCP) usage
- AI agents refine SQL queries
- Correlate metrics and logs
- Iterative diagnostic processes

What is the Model Context Protocol?

UNDERSTANDING MCP'S ROLE IN AI INCIDENT RESPONSE

The Model Context Protocol (MCP) provides a **standardized interface** enabling AI agents to introspect telemetry schemas and execute queries against historical datasets safely and efficiently.



AI Agent Workflow

NATURAL-LANGUAGE INPUT

The AI agent begins by receiving a **natural-language description** of the incident, allowing it to understand the context and specifics of the situation at hand.

SQL QUERY GENERATION

Based on the incident description, the agent generates **investigative SQL queries** tailored to extract relevant data from the telemetry sources, streamlining the data retrieval process.

ITERATIVE ANALYSIS

Once queries are executed, the agent analyzes the results, iteratively refining the queries as necessary to correlate telemetry data across multiple services and enhance diagnostic accuracy.

The Importance of Context Engineering

SCHEMA DOCUMENTATION

Extensive schema documentation is vital for ensuring that AI models can accurately interpret data structures, enhancing the effectiveness of incident diagnostics in complex systems.

CURATED LIBRARIES

Curated query libraries streamline the process for engineers by providing reliable, pre-tested SQL queries that reduce the risk of errors during incident investigations and enhance efficiency.

ARCHITECTURE DIAGRAMS

Architecture diagrams offer a clear visual representation of system components, which facilitates quicker understanding of interactions and dependencies, ultimately supporting more effective incident response efforts.

Challenges of Telemetry

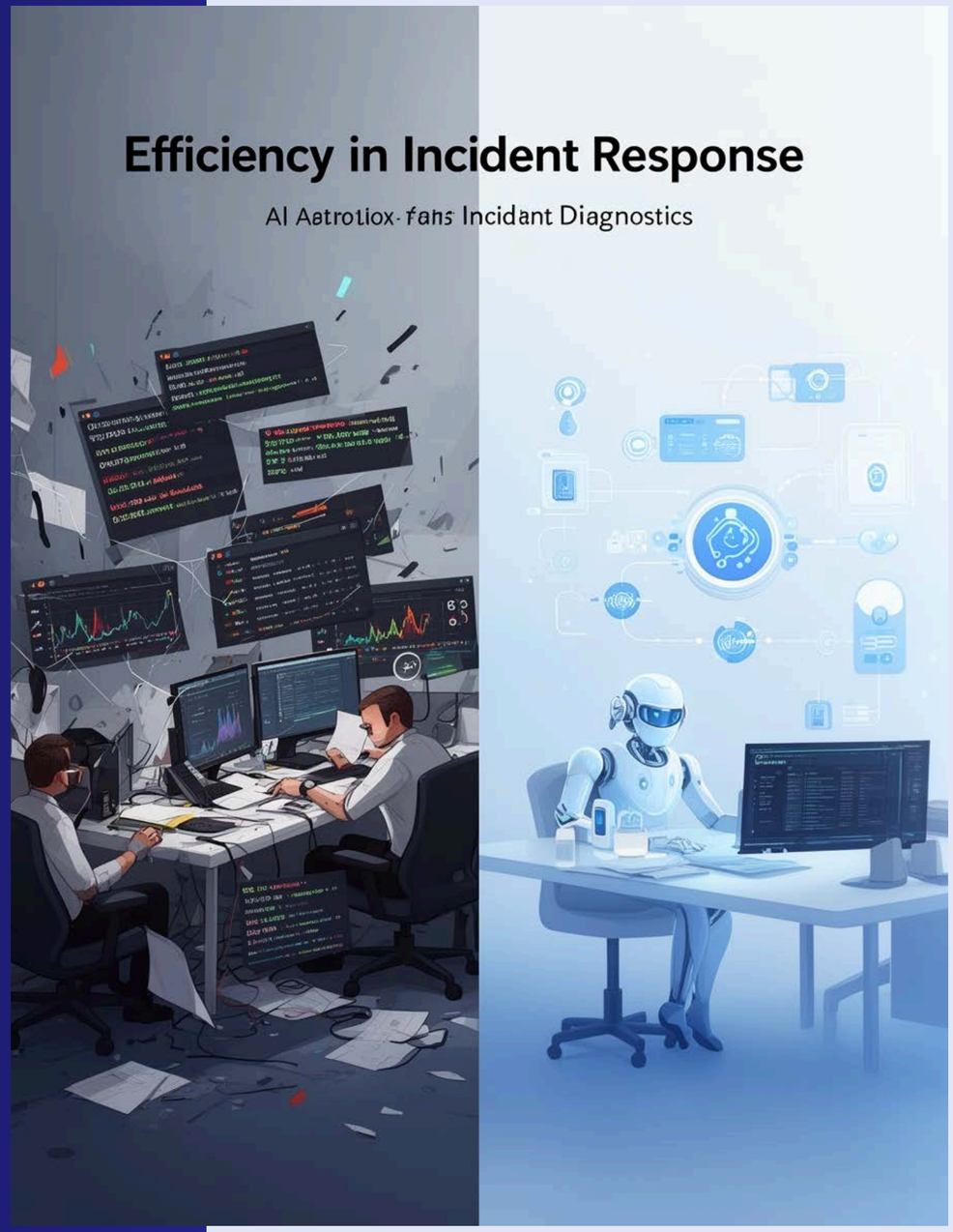
UNDERSTANDING COMPLEXITY IN DISTRIBUTED SYSTEMS

The **diversity and complexity** of telemetry across distributed systems leads to significant challenges, including heterogeneous schemas and varying timestamp skews, complicating incident diagnostics and resolution.Effective telemetry solutions require robust tools for data aggregation and normalization, ensuring that disparate data sources can be seamlessly integrated and analyzed. Additionally, implementing intelligent alerting mechanisms can help in proactively identifying anomalies before they escalate into critical issues. Collaboration between development and operations teams is essential to streamline processes and improve system resilience. Emphasizing a culture of continuous learning and adaptation can further aid in overcoming the hurdles posed by the intricate nature of telemetry in distributed environments, ultimately enhancing operational efficiency and system reliability.

Impact

“Reduced investigative effort by X%, Time to resolution shortened by Y%”

– CASE STUDY HIGHLIGHTS



Before AI Assistance: Manual Processes

- 2-4 hours spent on diagnostics
- Manual SQL query writing
- Backend experts bottlenecking incidents
- Difficulty in correlating data
- Long resolution times, impacting service reliability

After AI Assistance: Streamlined Operations

- Automated SQL query generation
- Faster iterative refinements of queries
- Enhanced data correlation across services
- Reduced time to resolution
- Empowered junior engineers to troubleshoot effectively

Architectural Patterns Overview

LLM INTEGRATION

Integrating Large Language Models (LLMs) allows for intelligent processing of incident data, enhancing diagnostics by automating query generation and analysis across distributed systems.

CONTEXT ENGINEERING

Context engineering establishes clear documentation and architecture diagrams, enabling precise AI model training and improving incident response effectiveness through well-defined schemas and query libraries.

HUMAN OVERSIGHT

Implementing human-in-the-loop mechanisms ensures that critical decisions benefit from expert review, balancing automation with the necessary oversight to maintain operational trust and accuracy.

Operational Lessons Learned

CURATED QUERY LIBRARIES

Maintaining **curated query libraries** is crucial for reliability in incident response. These libraries enhance the efficiency of AI systems by providing a solid foundation for generating accurate queries.

SCHEMA DOCUMENTATION

Comprehensive **schema documentation** is essential for effective AI model training. It enables agents to understand data structures better, facilitating more precise query generation and improved incident diagnostics.

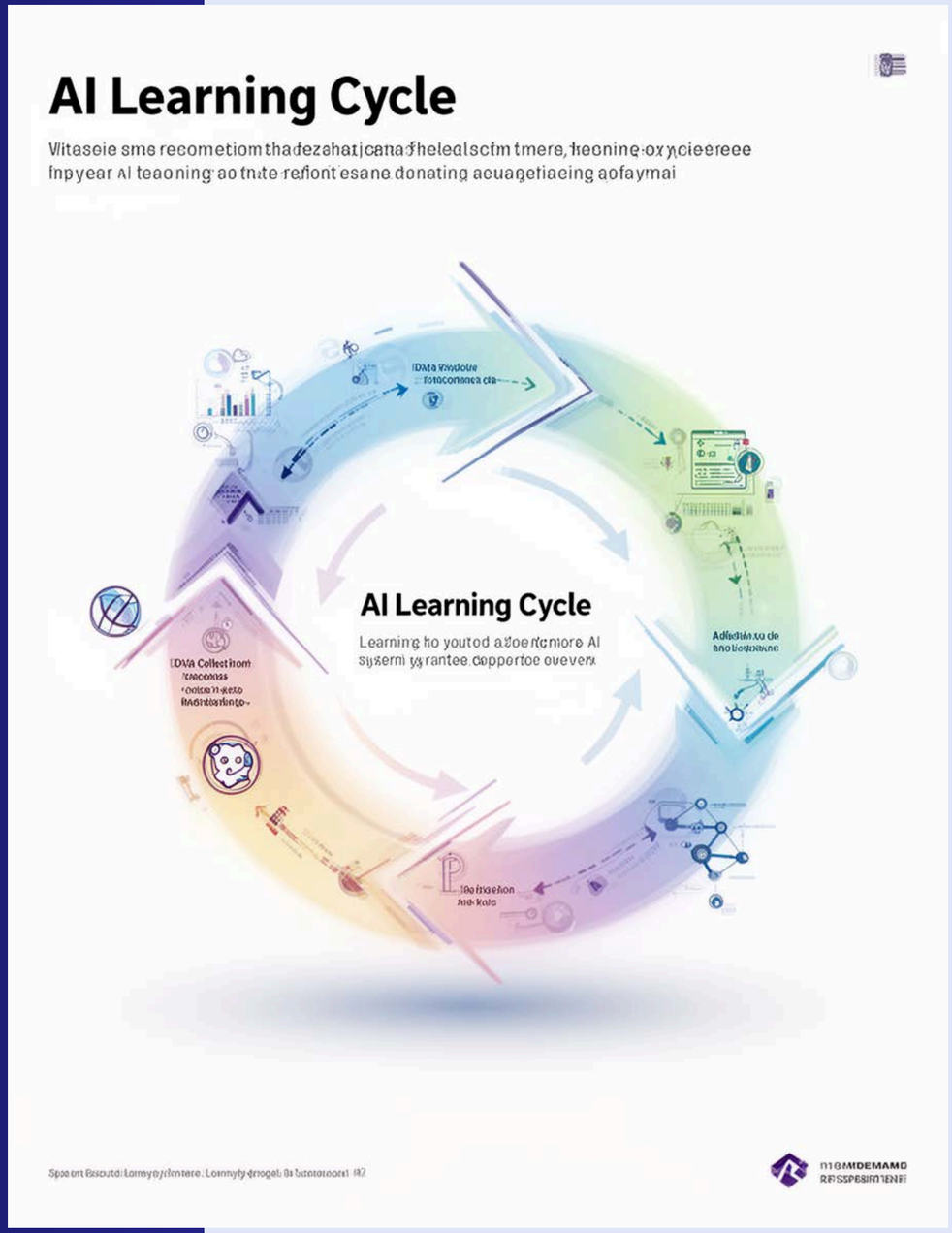
BALANCING AUTOMATION

Striking a balance between **AI automation** and human oversight is vital. This ensures that while efficiency is enhanced, crucial human judgment is preserved in complex incident analysis.

Future Directions

Enhancements in AI Assistance

- Expanding incident pattern catalogs
- Enhanced model context techniques
- Broader integration with tools
- Adaptive learning from incidents



Next Steps for Integration

ASSESS TELEMETRY MATURITY

Evaluate your current telemetry setup to identify gaps in data quality and schema consistency, ensuring that you can leverage AI-assisted diagnostics effectively and efficiently in your systems.

DEVELOP CONTEXT MATERIALS

Create comprehensive context engineering resources, including documentation, diagrams, and curated query libraries, to support the integration of AI into your incident response workflows and enhance reliability.

PILOT AI DIAGNOSTICS

Initiate trials of AI-assisted diagnostic tools within your team, using MCP-enabled interfaces to streamline incident investigations while maintaining human oversight for optimal accuracy and trustworthiness.

Thank you for your attention!

– Makarand Gujarathi