**PROMPT GENERATION**

As a prompt engineering expert, pls write a detailed prompt using 6-point magic prompt formulae with mark-down prompting technique comprising of: Persona, Context, Task, Example, Format and Tone, for the following problem.

I am an experienced lead data scientist working in cyber security data science/AI/ML Group for a large bank. We solve cyber sercurity problems and challenges by building advanced analytical solutions such as Gen AI applications, AI Chatbot UIs, AI Agents, and traditional predictive AI/ML Models and deploying them in production.

We use GCP Platform for building Gen AI solutions, Hadoop Data Lake for data integration and aggregation, MongoDB for UI Apps and Python for coding.

Since we are deploying such Gen AI/AI/ML products in production, we need to build and deploy a AI Ops framework and platform for effective 360 degree monitoring and measuring performance of all components of deployed solutions.

You are a AI Ops expert with deep knowledge of building, monitoring and measuring performance of various types of advanced analytics platforms and products.

Please recommend a comprehensive AI Ops strategy with short-term and long-term road map for building a successful and effective AI Ops monitoring platform for monitoring and measuring Gen AI Apps/ Chatbots / AI Custom LLMs / ML Models:

The strategy should have following features:

1. AI Ops Logical Architecture Diagram
2. AI Ops Physical Architecture Diagram
3. Metric Groups and metrics with definitions and calculation logic for different modules/components such as data integration pipelines, data, LLM/AI/LLM Model performance, Model Results, UI Experience etc
4. Short-term Roadmap
5. Long-term Roadmap

**GENERATED GENERAL PROMPT TEMPLATE**

**#\*\*Role:\*\***

You are an AI/ML technology educationalist and expert with deep knowledges of training programs being offered on various IT leading platforms such as PluralSight, Udemy, LinkedIn, Coursera etc.

I am an experienced IT Professional with 15 years of consulting and project/program management expertise, currently working at one of the top 5 consulting firms globally. My goal is to transition into an AI Generalist role by gaining foundational knowledge and hands-on experience in Natural Language Processing (NLP), Large Language Models (LLMs), and Prompt Engineering.

**#\*\*Objective:\*\***

Recommend top beginner-friendly online training programs, certifications, or learning paths that cover:

- 📘 NLP fundamentals and techniques

- 🤖 LLM architecture, training, and applications

- 🧵 Prompt engineering concepts and best practices

**#\*\*Instructions:\*\***

- Must be suitable for professionals with non-academic backgrounds in AI/ML

- Prioritize self-paced and flexible online learning platforms

- Include free or affordable options when possible

- Courses should offer practical projects or assignments

- Prefer resources that are up-to-date (2023+)

**#\*\*Output format:\*\***

Present a table with program name, provider, cost, estimated duration, and a brief description of focus areas.

**# AI Ops Strategy Prompt for 360-Degree Advanced Analytics Monitoring**

**## Persona**

You are an \*\*AI Ops expert\*\* with deep expertise in designing, building, and operating monitoring platforms for advanced analytics and AI/ML solutions in large-scale, cyber security-focused banking environments.

**## Context**

Your organization is a major bank with a data science AI/ML group that delivers Gen AI applications, AI chatbots and agents, and traditional predictive ML models. The technology stack includes \*\*GCP\*\* (Gen AI workloads), \*\*Hadoop Data Lake\*\* (integrations/pipelines), \*\*MongoDB\*\* (UI & analytics apps), and \*\*Python\*\* (coding).

As these systems are deployed in production, there is an urgent need for an enterprise-grade \*\*AI Ops framework\*\* for end-to-end monitoring, observability, and operational measurement across all components.

**## Task**

Recommend a comprehensive \*\*AI Ops strategy\*\*, including both a short-term and long-term roadmap, for building an effective 360-degree monitoring platform. Your deliverable must address:

- Logical architecture diagram (text/mermaid/markdown-based)

- Physical architecture diagram (text/mermaid/markdown-based)

- Clear definitions and calculation logic for key metric groups: data integration, data quality, AI/LLM/model performance, model outputs, UI/UX, and security/privacy

- Actionable short-term roadmap

- Scalable, automation-oriented long-term roadmap

- Explicit links between metric groups and business, security, and tech personas

**## Example**

\*For reference, in the metric groups table, you might define “Pipeline SLA Compliance” as:

`% of successful pipeline runs completed within target time window = 100\*(successful\_runs\_within\_SLA/total\_runs)`.\*

\*Diagrams can use Markdown-style ASCII art or mermaid.js for clarity, e.g.:\*

**## Format**

- Headings for each main deliverable

- Bullet points/tables for lists and metrics

- visual diagrams for architectures

- Tables for metric definitions

- Concise, clearly separated short-term and long-term roadmaps

**## Tone**

- Professional and \*\*solution-oriented\*\*

- Clear, actionable, and tailored for enterprise/banking context

- Avoid jargon unless essential for precision; define technical terms on first use

**RESPONSE**

**Comprehensive AI Ops Strategy for 360-Degree Monitoring of Gen AI, Chatbots, AI/LLM Models, and ML Solutions**

As an expert in AI Ops, here is a strategic, actionable, and bank-ready approach to establishing a robust AI Ops platform for end-to-end monitoring and performance measurement of all deployed advanced analytics and AI/ML services.

**1. AI Ops Logical Architecture Diagram**

Logical Layer Diagram: Key Components & Flow

* Observation Layer
  + Data/Log Ingestion (structured, semi-structured, unstructured)
  + Real-time and batch metrics collectors
* Analytics & Intelligence Layer
  + Metrics processing, anomaly detection, trend analysis
  + Root cause & impact analysis
  + Automated incident detection and predictive analytics
* Action Layer
  + Automated responses, remediation triggers
  + Notification and escalation engine
* Management & Governance Layer
  + Policy, audit, and security controls
  + Metric definitions, SLA/KPI management, reporting
* User Layer
  + Unified dashboards for operators, data scientists, InfoSec, and business users
  + Alert management, insights, and visualization

This logical separation ensures modularity, agility, and silo-breaking monitoring across all AI/ML \assets

**2. AI Ops Physical Architecture Diagram**

Physical Platform Stack: Technologies & Interfaces

* Data Sources:
  + GCP AI/ML/GenAI workloads
  + Hadoop Data Lake for pipelines
  + MongoDB-powered UI applications
* Data Collection & Integration:
  + Agentless log collection, API connectors, DB hooks for Hadoop, GCP, MongoDB
  + Custom metric exporters for Python-based models/apps
  + Streaming data buses (Kafka/PubSub)
* Processing & Storage:
  + Stream and batch processing engines (Spark, Dataflow)
  + Time-series metrics storage (e.g., Prometheus, BigQuery)
* AI/ML Ops Engines:
  + Central AIOps/observability tool (e.g., Datadog, Dynatrace, custom AI Engine)
  + Inbuilt AI/ML & rules-based analysis
* User Interface/Dashboards:
  + Unified dashboards (Grafana, Kibana, Looker)
  + Self-service alerting and visualizations mapped to business, security, and IT needs

The architecture is cloud-agnostic but leverages native GCP tools for tight integration with the bank’s stack, ensuring scalable monitoring[2](https://www.linkedin.com/pulse/ai-ops-platform-breakdown-its-architecture-john-benito-s9ute)[4](https://uptimerobot.com/knowledge-hub/monitoring/ai-monitoring-guide/)[5](https://learn.microsoft.com/en-us/azure/architecture/ai-ml/guide/machine-learning-operations-v2)[6](https://www.dynatrace.com/news/blog/how-to-implement-an-aiops-strategy-at-scale/).

**3. Metric Groups & Definitions**

| Metric Group | Example Metrics | Definition/Calculation Logic |
| --- | --- | --- |
| Data Integration | Pipeline latency, throughput, error rate, SLA compliance | (total records processed) / time taken, % failed jobs, SLA adherence rates |
| Data Quality | Missing values, schema drift, duplication, field-level profiling | % missing data, detection of schema changes, unique vs. duplicate records |
| LLM/AI Model Performance | Inference latency, accuracy, hallucination/error rate, prompt injection rate | Avg. response time, model-specific accuracy or F1, hallucination % |
| Model Output Quality | Drift metrics, bias detection, output stability, compliance events | Population stability index, bias score, regulatory compliance flags |
| Infrastructure | CPU/RAM/GPU utilization, disk I/O, network latency, process health | % resource utilization, health-check pass/fail counts, packet loss/latency |
| UI/UX Monitoring | Page load time, API error rates, user engagement, satisfaction/NPS | Avg. page load (ms), API 4xx/5xx error rates, survey NPS, session length |
| Security & Privacy | Auth failures, data leakage, prompt injection, PII exposure events | # unauthorized access, leakage incidents, real-time detection of attacks |

* All metrics should be aggregated, trended, and available for drill-down visualizations.
* Calculation logic should be automated with business-context enrichment for actionable insights[7](https://coralogix.com/ai-blog/evaluation-metrics-for-ai-observability/)[4](https://uptimerobot.com/knowledge-hub/monitoring/ai-monitoring-guide/).

**4. Short-Term Roadmap (0–9 months)**

* Assessment & Requirements: Map all deployed AI/ML/GenAI components, catalog metric needs, regulatory and business requirements.
* Tool Selection & PoC: Evaluate, pilot, and select AIOps, observability, and metrics tools compatible with Hadoop, MongoDB, GCP, and Python.
* Initial Platform Setup:
  + Integrate data sources and instrument metric collection.
  + Establish basic dashboards, health checks, and alerting for all key modules (pipelines, models, UIs).
* Metric Definition & Baselines:
  + Formalize metric groups, thresholds, SLAs.
  + Kick-off baseline benchmarking and alert tuning.
* Security & Compliance Controls: Implement access, encryption, anonymization, audit trails for monitored data.
* Quick Wins: Enable auto-remediation for known error patterns in pipelines and simple UI outages.
* Stakeholder Enablement: Train teams, establish clear runbooks and response protocols, launch initial reporting.

**5. Long-Term Roadmap**

* Automated AI-driven Root Cause & Remediation: Expand machine learning for anomaly detection, predictive insights, and auto-resolution across the stack (especially for LLM/GenAI-specific incidents).
* Self-healing Workflows: Enable self-orchestration for infra and application faults, reducing MTTR and downtime.
* Unified, Persona-Based Dashboards: Granular views tailored for security, business, and technical personas.
* Advanced Security Monitoring: Implement adversarial prompt detection, hallucination containment for Gen AI, and regulatory compliance surveillance[7](https://coralogix.com/ai-blog/evaluation-metrics-for-ai-observability/).
* Continuous Improvement: Evolve platform with new metric groups, support new tech stacks, regular reviews and innovation workshops.
* Full Lifecycle CI/CD Integration: Tighten together observability, AIOps, and MLOps pipelines for rapid safe deployments and rapid rollback.
* Cross-Domain Analytics: Blend AI Ops metrics with business and risk data for holistic operational risk reporting.
* Scalability: Cloud-native scaling—supporting hyper-growth, large model deployments, and new application launches smoothly.

**Best Practices:**

* Start small, iterate rapidly, automate everything possible.
* Balance real-time and historical views for both quick response and long-term learning.
* Continuously tune metrics and thresholds to reduce alert fatigue.
* Foster tight collaboration between data science, security, infrastructure, and business teams.

**By combining strong architecture, robust metric engineering, and staged delivery with automation and AI, your AI Ops platform will deliver the 360-degree monitoring essential for secure, efficient, and compliant AI-driven bankingnking operations.**

Below are text-based logical and physical architecture diagrams for the recommended AI Ops platform targeting monitoring/observability of Gen AI, LLM, ML, chatbot, and supporting modules within a cybersecurity-focused banking environment.

**Logical Architecture Diagram (Text)**

**┌────────────────────┐**

**│ User Layer │**

**│ (Dashboards, Alerts, Security Audits) │**

**└──▲─-------─────────────┘**

**│ │**

**┌───────────────┴───────┐ ┌───┴────────────────────────────┐**

**│ Management & │ │ Action Layer │**

**│ Governance Layer │ │(Auto-remediation, Alert Mgmt) │**

**│ (Policy, SLA, Security│ └────────────┬───────────────────┘**

**│ Controls) │ │**

**└──────────▲────────────┘ ┌────────┴──────────┐**

**│ │ Analytics & │**

**│ │ Intelligence Layer│**

**│ │(Anomaly detect, │**

**│ │ RCA, prediction) │**

**│ └─────▲─────────────┘**

**│ │**

**┌──────┴────────────┐ ┌──────┴─────────┐**

**│ Observation Layer │ │ Data Sources │**

**│ (Metrics/Log │<────│ (GCP, Hadoop, │**

**│ Collection & │ │ MongoDB, etc.) │**

**│ Enrichment) │ └───────────────┘**

**└───────────────────┘**

**Key Flows:**

* Data & logs flow bottom-up—from applications and pipelines to the Observation Layer.
* Analytics & action modules process telemetry for incident detection, alerting, and auto-remediation.
* Management provides governance and reporting.
* Operators/users interact with the top User Layer for insights and manual overrides[1](https://cloud.ibm.com/docs/industry-ref-arch?topic=industry-ref-arch-banking-app)[2](https://www.linkedin.com/pulse/aiops-architecture-what-how-changing-sharon-abraham-ratna)[5](https://www.linkedin.com/pulse/from-vision-execution-how-ai-first-thinking-shaping-future-silva-g8n6e).

**Physical Architecture Diagram (Text)**

┌──────────────────────────────────────────────────────┐

│ Unified UI Dashboards (Grafana, Kibana) │

└────────────────┬────────────────────────────────────┘

│

┌───────────┴────────────────────────┐

│ Central AIOps/Observability Tool │

│ (Datadog, Dynatrace, or Custom) │

└────────▲───────────▲──────────────┘

│ │

┌──────────┴───┐ ┌───┴──────────┐

│ Metrics Store│ │ Event Stream │

│(Prometheus, │ │(Kafka/PubSub) │

│BigQuery) │ └───▲───────────┘

└──────────────┘ │

▲ ▲

┌────┴─────┐ ┌─────┴───────────────┐

│ Agents & │ │ Data Collection │

│ Exporters│ │ & Integration Layer │

└────▲─────┘ │ (Log/API/Custom) │

│ └─────────▲-----------┘

┌────┴───────┐ ┌─────┴─────────────┐

│ GCP AI │ │ Hadoop Data Lake │

│ GenAI/LLMs │ └───────────────────┘

├────────────┤ ┌──────────────┐

│ MongoDB │<-----│ Python Apps │

└────────────┘ └──────────────┘

**Technology Stack:**

* Collection: Fluentd, API connectors, DB hooks for Hadoop/GCP/MongoDB.
* Processing: Spark, Dataflow (batch/stream), event/event mesh.
* Storage: Prometheus (metrics), BigQuery (logs).
* Core AIOps: Selected enterprise tool, or custom ML/inference/alerting pipelines.
* UI: Grafana/Kibana/Looker, connected to business, security, and operational users.
* Security: Integrated authentication, data masking, audit logging[1](https://cloud.ibm.com/docs/industry-ref-arch?topic=industry-ref-arch-banking-app)[2](https://www.linkedin.com/pulse/aiops-architecture-what-how-changing-sharon-abraham-ratna)[3](https://pmc.ncbi.nlm.nih.gov/articles/PMC9328990/)[5](https://www.linkedin.com/pulse/from-vision-execution-how-ai-first-thinking-shaping-future-silva-g8n6e).

**Diagram Notes:**

* These are textual depictions that map directly to recommended architecture domains and tech in your original strategy.
* If you require graphical (image-based) diagrams, please specify your desired format (e.g., a particular tool), and I can provide system-level pseudographics, mermaid.js, or detailed draw.io/XML export compatible with diagramming tools.
* <https://cloud.ibm.com/docs/industry-ref-arch?topic=industry-ref-arch-banking-app>
* <https://www.linkedin.com/pulse/aiops-architecture-what-how-changing-sharon-abraham-ratna>
* <https://pmc.ncbi.nlm.nih.gov/articles/PMC9328990/>
* <https://blog.lumen.com/modernizing-it-operations-with-aiops-a-comprehensive-guide/>
* <https://www.linkedin.com/pulse/from-vision-execution-how-ai-first-thinking-shaping-future-silva-g8n6e>
* <https://www.logicmonitor.com/blog/what-is-aiops>
* <https://docs.broadcom.com/docs/aiops-essentials-of-root-cause-analytics>
* <https://inoxoft.com/blog/10-requirements-for-building-digital-banking-architecture/>
* <https://learn.microsoft.com/en-us/azure/architecture/ai-ml/>
* <https://www.mckinsey.com/industries/financial-services/our-insights/platform-operating-model-for-the-ai-bank-of-the-future>