JUPITER PRIME

Consulting • Training • Advisory

جوبيتر برايم للاستشارات والتدريب

COURSE BROCHURE

ChatGPT Performance Monitoring and Optimisation Training Course

16 – 20 May 2027
Kuala Lumpur

ChatGPT Performance Monitoring and Optimisation Training Course

Ref: 61_1536 **Date:** 16 – 20 May 2027 **Location:** Kuala Lumpur **Fees:** 4900 GBP

ChatGPT Performance Monitoring and Optimisation is a 1-week training course for AI product owners, operations managers, digital teams and business analysts that ends with a ChatGPT monitoring dashboard and optimisation plan for one live assistant. Once a ChatGPT deployment reaches staff or customers, answer quality, token spend and response time drift week by week unless someone measures them. Jupiter Prime Training runs this ChatGPT Performance Monitoring and Optimisation course for teams operating assistants.

Why ChatGPT Performance Monitoring and Optimisation Matters Now

Three forces push ChatGPT performance monitoring and optimisation onto operating agendas: assistants moving from trials into help desks, sales and internal knowledge search; finance teams questioning usage bills that grow with every new user; and users who abandon an assistant after a few wrong or slow answers. An organisation that tracks quality, cost and latency can tune prompts, retrieval and model choice with evidence; one that does not loses trust and budget at once. The course draws on ISO/IEC 42001 for AI management and the NIST AI Risk Management Framework NIST AI RMF for measuring and managing assistant risk.

Who Gains Most from ChatGPT Performance Monitoring and Optimisation

- Product owners accountable for an assistant's usefulness, adoption and roadmap
- Operations leads who keep assistants available, responsive and within budget
- Digital specialists who write prompts, configure retrieval and maintain knowledge sources
- Analysts who build usage reports and translate feedback into improvement items
- Service and quality owners who sign off assistant answers in customer-facing channels

A digital or customer service function gains most when its product, operations and knowledge members attend together and agree one scorecard for the assistant.

ChatGPT Performance Monitoring and Optimisation Capabilities

Participants learn to treat a ChatGPT assistant like any production service: measured, tuned and reported with clear thresholds.

- Specify quality metrics such as groundedness, accuracy and refusal rate — product owners
- Build golden evaluation sets and regression tests for prompts — knowledge and digital teams
- Calculate cost per conversation and token budgets per use case — finance partners
- Test prompt, retrieval and model changes through controlled comparisons — release approvers
- Design user feedback loops that turn ratings into fixes — service leaders
-

Communicate assistant health through a dashboard with alert thresholds — operations and leadership

Key ChatGPT Performance Monitoring and Optimisation Terms

- Golden evaluation set — a fixed list of representative questions with approved answers used to score an assistant before and after changes.
- Groundedness — the degree to which an answer is supported by the source documents the assistant retrieved.
- Hallucination rate — the share of answers containing claims that are false or unsupported by any source.
- Token — a small unit of text the model reads or writes, used to measure context length and usage cost.
- Latency — the time between a user's question and the assistant's complete answer.
- Retrieval-augmented generation — a pattern where the assistant searches approved documents and answers from them rather than from memory alone.

ChatGPT Performance Monitoring and Optimisation Day by Day

Day 1: Assistants in Production

- Assistant Use-Case Inventory — Help Desk, Sales and Internal Search
- Service Level Objectives for Assistants — Quality, Speed and Availability
- Failure Mode Catalogue — Hallucination, Refusal, Drift and Outage
- Banking and Retail Assistant Cases — Lessons From Live Deployments
- Stakeholder Scorecard Design — What Each Sponsor Needs to See

Day 2: Measuring Answer Quality

- Golden Evaluation Set Construction — Representative Questions and Approved Answers
- Groundedness and Accuracy Scoring — Human Raters Versus Model Graders
- Hallucination and Refusal Rate Tracking — Release Thresholds
- Conversation Log Sampling — Privacy-Safe Review of Real Use
- Inter-Rater Agreement Checks — Trust in Quality Scores

Day 3: Prompt and Retrieval Optimisation

- System Prompt Versioning — Controlled Changes and Rollback
- Retrieval-Augmented Generation Tuning — Chunking, Ranking and Source Freshness
- A/B and Shadow Testing — Evidence Before Promoting a Change
- Regression Test Suites — Protecting Answers That Already Work
- Model and Context Window Selection — Quality Against Cost

Day 4: Cost, Latency and Risk Controls

- Token Budgeting per Use Case — Cost per Conversation Targets
- Caching and Response Streaming — Latency Reduction Options
- Usage Anomaly Alerts — Spikes, Abuse and Runaway Spend
- NIST AI RMF Measure and Manage — Assistant Risk Thresholds
- ISO/IEC 42001 Monitoring Records — Evidence of Ongoing Control

Day 5: Practising the Monitoring Dashboard and Optimisation Plan

- Service Objectives From Day 1 — Dashboard Layout and Owners
- Quality Metrics From Day 2 — Weekly Evaluation Routine
- Optimisation Tests From Day 3 — Prioritised Change Backlog
- User Feedback Loop Design — Ratings, Comments and Fix Tracking
- ChatGPT Monitoring Dashboard and Optimisation Plan — Sponsor Sign-Off Pack

Applying ChatGPT Performance Monitoring and Optimisation After the Course

- Assemble a golden evaluation set of fifty real questions for one live assistant and score the current version for groundedness and hallucination rate.
- Calculate cost per conversation for the top three use cases and set token budgets with alert thresholds in the usage console.

ChatGPT Performance Monitoring and Optimisation Questions FAQ

What is ChatGPT performance monitoring and optimisation and why does it matter now?

ChatGPT performance monitoring and optimisation is the routine measurement and tuning of an assistant's quality, cost and speed. It matters now because assistants are moving into daily operations, where wrong answers and rising usage bills quickly erode trust and budget.

Which roles gain most from ChatGPT performance monitoring and optimisation training?

AI product owners, operations managers, digital teams and business analysts gain most. Service owners who approve customer-facing answers also benefit, since they set the quality thresholds the assistant must meet.

How does ChatGPT performance monitoring and optimisation differ from a prompt writing course?

It covers the whole operating loop rather than writing single prompts. You build evaluation sets, run controlled tests, track cost and latency and manage feedback, so changes are proven before release.

What do you bring back from ChatGPT performance monitoring and optimisation training?

You bring back a ChatGPT monitoring dashboard and optimisation plan for one live assistant. It defines metrics, thresholds, owners, a weekly evaluation routine and a prioritised change backlog.

ChatGPT Performance Monitoring and Optimisation: The Way Forward

The ChatGPT monitoring dashboard and optimisation plan informs the first operating decision: which quality or cost gap to fix in the next release. Product owners, operations leads and analysts use it in weekly reviews. After a month of live readings, revisit the golden evaluation set and token budgets, since real questions and usage patterns shift as adoption grows.