informatech

DIGITAL INNOVATION AND TRANSFORMATION | DIT-023

ChatGPT: Performance Monitoring and Optimization

UK
+44 33 000 111 90
info@informatech.co.uk
https://informatech.uk
63-66 Hatton Garden Hatton Garden
EC1N 8LE , London

NL
+31 85 74 444 46
info@infomatech.nl
https://infomatech.nl
Waarderweg 50 - 2031PB
Haarlem - Netherlands

Course content

Why Attend

As organisations increasingly integrate ChatGPT and generative artificial intelligence into business operations, customer services, knowledge management, and digital workflows, ensuring consistent performance has become a critical management responsibility. Successful AI adoption requires more than deploying a functional application; organisations must continuously evaluate response quality, operational efficiency, reliability, security, and business value.

AI-powered systems can experience performance challenges involving inaccurate responses, inconsistent outputs, excessive latency, high operational costs, inefficient prompts, and changing user requirements. Applications that perform well during initial testing may become less effective as business information, usage patterns, model versions, and operational conditions evolve.

Effective performance monitoring enables organisations to understand how AI applications behave in real-world environments, identify weaknesses, measure user satisfaction, and implement improvements based on reliable evidence. Performance optimisation helps improve response quality, reduce unnecessary processing, manage costs, and strengthen the overall user experience.

The ChatGPT: Performance Monitoring and Optimization course provides participants with a practical framework for evaluating, monitoring, and improving ChatGPT-based applications and workflows.

Participants explore AI performance indicators, prompt evaluation, response accuracy, latency measurement, token usage, cost optimisation, application observability, quality assurance, model selection, and continuous improvement.

The programme also examines AI governance, monitoring dashboards, testing methodologies, human feedback mechanisms, and the management of performance risks associated with generative AI systems.

Through practical workshops, analytical exercises, and implementation case studies, participants develop the skills needed to establish effective monitoring processes and optimise ChatGPT solutions for reliable, efficient, and responsible organisational use.

Professional Problems This Course Is Designed to Solve

• Inconsistent AI response quality: Establishing systematic evaluation methods to assess accuracy, relevance, completeness, consistency, and instruction adherence.

• Slow response times and poor user experience: Identifying latency bottlenecks and applying practical optimisation techniques to improve application responsiveness.

Course content

Why Attend

• Uncontrolled AI operating costs: Monitoring token consumption, model usage, request volumes, and processing costs to improve resource efficiency.

• Limited visibility into AI application performance: Developing monitoring dashboards and operational indicators that provide meaningful information about system behaviour.

• Difficulty identifying AI errors and performance deterioration: Implementing testing, feedback collection, exception monitoring, and regression evaluation procedures.

• Inefficient prompts and workflow configurations: Improving prompt structure, context management, model selection, and application architecture.

• Weak AI governance and accountability: Establishing performance standards, monitoring responsibilities, audit trails, and continuous improvement processes.

Course Methodology

The programme combines instructor-led discussions, live demonstrations, AI performance analysis exercises, prompt optimisation workshops, application monitoring simulations, and practical case studies.

Participants work with illustrative ChatGPT application scenarios involving customer support, document processing, internal knowledge assistance, business analysis, and workflow automation.

Hands-on activities include developing performance indicators, evaluating AI-generated responses, comparing prompt variations, analysing latency and token usage, designing monitoring dashboards, and preparing optimisation recommendations.

The programme addresses both ChatGPT-based business workflows and API-integrated applications. Technical exercises may use illustrative application logs, evaluation datasets, and monitoring tools, while non-technical participants focus on performance interpretation, quality management, and operational decision-making.

Course Objectives

By the end of this course, participants will be able to:

Course content

Course Objectives

- Explain the principles of performance monitoring for ChatGPT and generative AI applications
- Identify relevant AI quality, reliability, efficiency, and business performance indicators
- Develop structured methods for evaluating AI-generated responses
- Measure and interpret latency, throughput, token usage, and operational costs
- Apply prompt optimisation techniques to improve response quality and efficiency
- Evaluate model selection and configuration decisions against business requirements
- Design monitoring dashboards for AI application performance
- Identify errors, regressions, and changes in AI system behaviour
- Establish testing and evaluation procedures for AI workflows
- Apply security, privacy, and governance controls to AI performance monitoring
- Develop continuous improvement processes based on operational evidence and user feedback
- Prepare an integrated ChatGPT performance monitoring and optimisation framework

Target Audience

This course is suitable for:

- Artificial intelligence and machine learning professionals
- Digital transformation managers and specialists
- IT operations and application support managers
- Software developers and AI application engineers
- Data analysts and business intelligence professionals
- AI product managers and digital service owners
- Business process improvement professionals
- Customer experience and service automation managers

Course content

Target Audience

• Technology governance and AI risk professionals

• Project managers responsible for generative AI implementation

Target Competencies

Participants will develop competencies in:

• Generative AI performance measurement

• AI response quality evaluation

• Prompt engineering and optimisation

• AI latency and throughput analysis

• Token usage and cost management

• AI application monitoring and observability

• Model evaluation and regression testing

• Performance dashboard development

• AI operational risk and governance

• Continuous AI performance improvement

Course outline

Day 1: ChatGPT Performance Fundamentals and Monitoring Frameworks

• Understanding performance management in generative AI applications and the relationship between response quality, operational efficiency, reliability, and business outcomes

• Identifying common ChatGPT performance challenges, including hallucinations, inconsistent responses, excessive latency, context limitations, and inefficient processing

Course content

Course outline

• Establishing performance measurement frameworks covering quality, accuracy, relevance, response time, availability, cost, and user satisfaction

• Distinguishing performance considerations for direct ChatGPT usage, custom AI workflows, and API-integrated applications

• Defining performance baselines, service objectives, evaluation criteria, reporting responsibilities, and monitoring requirements

• Practical workshop: Developing a ChatGPT performance measurement framework with KPIs, evaluation criteria, target thresholds, and management reporting requirements

Day 2: AI Response Quality Evaluation and Prompt Optimisation

• Evaluating AI-generated responses for factual accuracy, completeness, relevance, clarity, consistency, and adherence to instructions

• Developing representative evaluation datasets and test scenarios that reflect real organisational tasks and user expectations

• Applying prompt optimisation techniques involving clear instructions, contextual information, examples, structured outputs, and task decomposition

• Comparing prompt variations through controlled testing and assessing improvements in response quality and token efficiency

• Understanding automated evaluation methods, human review, scoring rubrics, and the limitations of AI-assisted response assessment

• Hands-on exercise: Evaluating multiple prompt configurations and developing an improved prompt template supported by measurable quality indicators

Day 3: Latency, Token Usage, Cost Management and Technical Optimisation

• Understanding response latency, request processing time, throughput, and factors affecting AI application responsiveness

• Analysing token consumption, context size, output length, model selection, and their influence on application costs and performance

Course content

Course outline

• Applying optimisation techniques such as prompt simplification, context reduction, appropriate model selection, caching where supported, and efficient workflow design

• Evaluating API usage patterns, rate limits, concurrency considerations, retry mechanisms, and error handling in AI-enabled applications

• Comparing performance trade-offs involving speed, accuracy, reliability, model capability, and operating expenditure

• Practical workshop: Analysing illustrative AI application usage data and preparing a cost and latency optimisation plan with measurable performance targets

Day 4: AI Application Monitoring, Observability and Performance Dashboards

• Understanding AI application observability through operational logs, request metrics, response evaluation results, error records, and user feedback

• Designing monitoring dashboards covering request volumes, response latency, token usage, estimated costs, quality scores, and system reliability

• Identifying performance anomalies, recurring failures, unusual usage patterns, and changes in application behaviour

• Establishing alert thresholds, escalation procedures, incident investigation processes, and performance reporting cycles

• Applying regression testing to detect performance changes following prompt modifications, model updates, knowledge source changes, or workflow adjustments

• Case study: Developing an AI performance dashboard and investigating a simulated deterioration in response quality, latency, and operational efficiency

Day 5: AI Governance, Continuous Improvement and Performance Optimisation Strategy

• Establishing AI performance governance covering accountability, evaluation standards, documentation, monitoring responsibilities, and management oversight

• Evaluating privacy, security, data retention, access controls, and responsible AI considerations associated with application monitoring and operational logs

Course content

Course outline

• Integrating user feedback, human evaluation, incident findings, and performance data into structured continuous improvement processes

• Developing AI testing and optimisation cycles covering baseline assessment, improvement hypotheses, controlled changes, validation, and post-deployment monitoring

• Preparing an AI performance improvement roadmap that prioritises quality, reliability, cost efficiency, user experience, and measurable business benefits

• Capstone exercise: Developing and presenting a Comprehensive ChatGPT Performance Monitoring and Optimisation Framework incorporating performance KPIs, evaluation datasets, prompt improvement techniques, latency and cost analysis, monitoring dashboards, governance controls, and a continuous improvement roadmap

Seminar dates

Available seminar dates

Live dates and pricing for ChatGPT: Performance Monitoring and Optimization generated from the course details page.

Date	Location	Format	Fee
Dates on request	Venue on request	Classroom	**Contact us**

Live online option	Online delivery is available at €1,850.-.