Testing Generative AI Systems Training

Duration: **1 Day**

BILGINC IT Academy
info@bilginc.com
bilginc.com

Testing Generative AI Systems is a practical course designed for professionals who need to test, evaluate, and govern generative AI systems under real-world conditions.

Generative AI solutions behave differently from traditional software systems because the same input does not always produce exactly the same output. As a result, conventional deterministic testing approaches may not be sufficient on their own to assess quality, safety, and behaviour. This course helps participants understand that difference and develop testing methods suited to probabilistic, variable, and evolving AI systems.

Throughout the programme, learners work with realistic examples, structured testing frameworks, red-teaming exercises, and hands-on evaluation activities. Participants examine why generative AI systems fail, how risks such as hallucination, bias, toxicity, privacy issues, and drift emerge, and how those risks can change after deployment.

By the end of the course, participants will have a practical approach for evaluating GenAI systems throughout the product lifecycle rather than treating testing as a one-off activity.

What Participants Will Take Away

By the end of the day, participants will have developed:

- A practical testing strategy for a real generative AI use case,
- A repeatable framework for evaluating quality, risk, and behaviour in GenAI systems,
- Hands-on experience with red teaming and non-deterministic testing,
- Clear guidance for integrating GenAI testing into product and delivery lifecycles.

Prerequisites

Participants should have:

- Basic familiarity with digital products, AI, or software delivery,
- Some exposure to AI-enabled features or systems as a user, builder, or stakeholder.

No previous experience in AI testing, data science, or machine learning is required.

Testing Generative AI Systems Training

Duration: **1 Day**

BILGINC IT Academy
info@bilginc.com
bilginc.com

Who Should Attend

This course is designed for:

- QA Engineers and Test Leads responsible for AI-enabled systems,
- Product Managers and Product Owners introducing generative AI features,
- Developers and AI Practitioners building or integrating GenAI models and services,
- Risk, Compliance, and Governance professionals overseeing AI usage,
- UX, CX, and Innovation teams focused on trust, safety, and reliability.

What You Will Learn

By completing this course, participants will be able to:

- Explain why traditional testing approaches can be insufficient for generative AI systems,
- Identify major GenAI risk categories including hallucinations, bias, toxicity, privacy, and drift,
- Design testing strategies for non-deterministic and evolving outputs,
- Define meaningful benchmarks and evaluation criteria for GenAI quality,
- Apply red-teaming techniques to uncover hidden and adversarial failures,
- Balance automation with human judgement in AI testing,
- Integrate GenAI testing into continuous delivery and governance practices.

Training Outline

Module 1: Why Generative AI Breaks Traditional Testing

This module explores the fundamental differences between conventional software testing and testing generative AI systems.

- Deterministic versus probabilistic behaviour
- The impact of variable outputs on testing
- Limitations of traditional pass/fail approaches
- Uncertainty and behavioural variation in AI systems
- Developing a different testing mindset for GenAI

Module 2: Understanding GenAI Risk and Behaviour

Participants examine the main risks associated with generative AI and how they may evolve over time.

- Hallucination risks
- Bias and fairness concerns
- Toxic or inappropriate outputs
- Privacy and sensitive data risks
- Model and behavioural drift
- How risks differ before and after deployment
- The effect of user behaviour on system risk

Module 3: Designing Tests for Non-Deterministic Systems

Testing Generative AI Systems Training

Duration: **1 Day**

BILGINC IT Academy
info@bilginc.com
bilginc.com

This section focuses on building structured testing approaches for systems that may produce different outputs for the same request.

- Define expected behaviour ranges instead of single fixed outcomes
- Build test scenarios and test datasets
- Use repeated testing
- Evaluate output consistency and variability
- Identify important edge cases
- Define areas where human review is necessary

Module 4: Benchmarking and Evaluation Criteria

Participants learn how to assess GenAI quality using a broader set of measures than technical correctness alone.

- Build suitable quality benchmarks
- Define evaluation criteria
- Assess accuracy, relevance, and usefulness
- Measure safety and reliability
- Combine human and automated evaluation
- Create acceptance criteria for different use cases

Module 5: Red Teaming and Adversarial Testing

This module introduces approaches for testing how systems behave when faced with unexpected, manipulative, or hostile inputs.

- Introduction to red teaming
- Create adversarial prompts and scenarios
- Surface hidden failure modes
- Test security and safety controls
- Evaluate guardrails
- Feed red-team findings into risk management

Module 6: Testing in Production and Governance

The final module treats GenAI testing as a continuous activity that continues after deployment.

- Connect production monitoring with testing
- Evaluate live system behaviour
- Monitor drift and emerging risks
- Apply continuous evaluation
- Feed testing results into governance and risk processes
- Integrate GenAI testing into product and delivery lifecycles
- Continuously strengthen trust, safety, and quality