

PRACTICAL AI ASSURANCE WORKSHOP

TEST AND EVALUATION OF LARGE LANGUAGE MODELS AND GENERATIVE AI APPLICATIONS



This practical half-day workshop is designed to provide participants with a practical understanding of the process of testing and evaluating large language models and related generative AI applications for use in a business context.

Session 1: Evaluating Large Language Models	Session 2: Practical Techniques and Best Practices in LLM Testing
<p>Part 1: Introduction to Large Language Models</p> <ul style="list-style-type: none">• Overview of LLMs: Definition, evolution, and significance.• Patterns for Gen AI applications.• Key characteristics of LLMs: Scale, capabilities, and common applications. <p>Part 2: Testing Large Language Models</p> <ul style="list-style-type: none">• The importance of testing in LLM development.• Types of tests for LLMs: Functional testing, performance testing, safety testing.• Test case design: Considerations for comprehensive and effective test cases. <p>Part 3: Evaluation Metrics and Techniques</p> <ul style="list-style-type: none">• Common metrics for evaluating LLMs.• Understanding qualitative evaluations: Human judgment, interpretability, and explainability.• Challenges in evaluating LLMs: Subjectivity, bias, and ethical considerations. <p>Part 4: Practical Exercise</p> <ul style="list-style-type: none">• Hands-on activity: Designing test cases for a given LLM scenario.• Group discussion: Sharing insights and strategies for effective LLM testing.	<p>Part 1: Advanced Testing Techniques</p> <ul style="list-style-type: none">• Techniques for uncovering biases and ensuring fairness in LLMs.• Testing for robustness: Adversarial testing, stress testing. <p>Part 2: Optimising LLMs Through Iterative Testing</p> <ul style="list-style-type: none">• The role of iterative development in refining LLMs.• Incorporating feedback loops: Continuous testing and refinement.• Case studies: Examples of iterative testing and optimisation in LLMs. <p>Part 3: Best Practices and Future Directions</p> <ul style="list-style-type: none">• Best practices in testing and evaluating LLMs.• Emerging trends and future challenges in LLM testing.• Ethical considerations and responsible AI in the context of LLMs. <p>Part 4: Wrap-up</p> <ul style="list-style-type: none">• Summary of key takeaways from the workshop.• Resources for further learning and exploration.