Safe and Secure Use of Large Language Models (LLMs)

How to avoid the pitfalls of generative AI in the age of heightened privacy and disinformation

Focus on: The Manager ☑ The Specialist ☑ Spotlight Programme ☐ Hands-on Skills ☑

Course Overview

In an era where generative AI is transforming industries, understanding how to use Large Language Models (LLMs) safely and securely is crucial. This five-day course provides delegates with a comprehensive guide to navigating the risks of AI, including privacy concerns, data security, misinformation, and ethical considerations. Through interactive sessions, real-world case studies, and hands-on exercises, participants will learn how to mitigate threats, ensure compliance with regulations such as GDPR, and implement secure AI practices in the workplace. By the end of the course, attendees will be equipped with the knowledge and tools to use AI responsibly, protect sensitive information, and prevent the spread of disinformation.

Course Objectives		This Course is Ideal For:	
1	Explain how Large Language Models work, their capabilities, and the key risks associated with their use, including privacy breaches, security vulnerabilities, and misinformation.	✓	Business & Compliance Leaders – Those responsible for data and AI governance, risk management, and regulatory compliance in their organisations.
2	Implement strategies to safeguard sensitive data, comply with legal frameworks (e.g., GDPR, AI Act), and minimise the risk of data exposure when interacting with AI systems.	✓	Cybersecurity & IT Professionals – Individuals who oversee data security, threat mitigation, and safe AI deployment.
3	Recognise how AI can amplify disinformation, assess AI-generated content for accuracy and bias, and apply fact-checking techniques to ensure responsible usage.	✓	Content & Communications Teams – Those using AI for content creation, fact-checking, and mitigating misinformation risks.
4	Establish workplace guidelines and governance frameworks for the ethical and secure deployment of AI tools, preventing unauthorised data sharing and ensuring compliance with industry standards.	✓	AI Developers & Data Scientists – Technical professionals designing, deploying, or auditing AI systems for security and fairness.
5	Anticipate future risks associated with AI advancements, engage in adversarial testing to identify vulnerabilities, and promote ethical decision-making in AI applications.		

Course Content

Day	Theme	Coverage
1	Introduction to LLMs and Their Risks	Session 1: How LLMs Work and Their Capabilities • Overview of Large Language Models (e.g., GPT, Claude, Gemini) • How LLMs are trained and generate responses

Safe and Secure Use of Large Language Models (LLMs)

How to avoid the pitfalls of generative AI in the age of heightened privacy and disinformation

Focus on: The Manager ☑ The Specialist ☑ Spotlight Programme ☐ Hands-on Skills ☑

		• Understanding prompting techniques and model behaviour Session 2: Risks of LLMs – Privacy, Security, and Misinformation • Common risks: data leaks, adversarial inputs, bias, and misinformation • AI misinformation incidents (e.g., deepfakes, fake news) • Introduction to ethical and responsible AI use
2	Privacy Concerns and Data Security	Session 1: Protecting Sensitive Data in AI Interactions • How LLMs process and store data – what happens to user inputs? • Risks of unintended data leakage in workplace AI use • Best practices for data anonymisation and safe AI queries Session 2: Compliance and Regulatory Frameworks for AI • Overview of GDPR and AI Act regulations • Legal implications of AI-generated content and data handling • Corporate responsibility: Ensuring AI compliance in the workplace
3	Mitigating AI-Generated Misinformation and Bias	Session 1: AI Misinformation – How It Spreads and How to Counter It • How AI models can hallucinate facts • The role of AI in spreading fake news and propaganda • Fact-checking techniques for AI-generated content Session 2: Bias in AI – Detection and Mitigation • Understanding algorithmic bias in LLMs • Real-world consequences of biased AI outputs • How to audit and evaluate AI-generated responses for fairness
4	Secure AI Usage in the Workplace	Session 1: Developing AI Usage Policies for Organisations • How to create AI governance frameworks for businesses • Guidelines for safe internal AI use (e.g., handling confidential data) • Preventing accidental data sharing with AI tools Session 2: Preventing AI-Powered Social Engineering & Cybersecurity Risks • Understanding AI-driven phishing, deepfakes, and fraud • Threat modelling: Recognising and mitigating adversarial AI use cases
5	Futureproofing Against AI Threats and Ethical Considerations	Session 1: Emerging AI Threats and Security Strategies • What's next? The evolving risks of AI in cybersecurity and disinformation • Testing AI models for vulnerabilities • The role of explainable AI (XAI) and transparent AI systems Session 2: Ethical AI Use and Responsible Deployment • Ensuring AI aligns with ethical standards and human values

Safe and Secure Use of Large Language Models (LLMs)

How to avoid the pitfalls of generative AI in the age of heightened privacy and disinformation

Focus on: The Manager ☑ The Specialist ☑ Spotlight Programme ☐ Hands-on Skills ☑

		• The role of AI safety, fairness, and accountability in deployment • The importance of human oversight in AI-assisted decision-making

Course Assessment	Certification
Participants will be assessed on:	Upon successful completion of the course, participants will receive a **Certificate of Successful Completion**, along with a **Transcript of Marks** showing the performance by grade in each element of assessment and overall.
Participation in sessions	
Completion of exercises & case studies	
Performance in assessments	

Course Instructor

The speaker for this programme is an International Expert and Professor of Cyber Security at a leading UK university. He is also a Member of the UK Government's Cyber Security Advisory Board.
The speaker has extensive expertise in Cyber Security and Artificial Intelligence and is widely published in this field.