COURSE SYLLABUS

AI Red Teaming and Assurance Course (Self-Paced)

Twelve hours of self-paced AI red teaming and assurance — engagement planning, adversarial testing technique, assurance cases, and testing agentic systems. No lab required.

> Group classes in Live Online and onsite training is available for this course. For more information, email onsite@graduateschool.edu or visit: https://www.graduateschool.edu/courses/ai-red-teaming-and-assurance-self-paced

support@graduateschool.edu •
(888) 744-4723

Course Outline

Module 1

- Foundations: What Is AI Red Teaming, and Where Does It Fit in AI Assurance?
- What red teaming is and is not
- Where it sits in the AI assurance lifecycle
- The range of assurance activities
- The guidance the course draws on
- Who is involved

Module 2

- Planning and Scoping a Red-Team Engagement
- The engagement lifecycle
- Scoping to risk classification
- Setting red-team objectives
- Rules of engagement
- Team composition and deconfliction

Module 3

- Adversarial Testing Techniques for Machine Learning and Generative AI
- Matching technique to attack surface
- Manual versus automated and tool-assisted red teaming
- Generative AI probing techniques
- A preview of testing agentic and tool-using systems
- Telling red-team findings apart from ordinary bugs

Module 4

- AI Assurance Frameworks: TEVV, Assurance Cases, and Documentation
- Test, Evaluation, Verification and Validation
- The assurance case as claim, argument and evidence
- Model cards and system cards
- Independent evaluation and measurement science
- From assurance documentation to a risk-acceptance decision

Module 5

- From Findings to Fixes: Reporting, Remediation, and Human Oversight
- Structuring a findings report
- Prioritizing by impact rather than technical complexity
- Remediation and retesting
- Human oversight as a mitigation category
- When a finding becomes an incident

Module 6

- Red Teaming and Assurance for Generative and Agentic AI; Capstone Scenario
- Foundation models and misuse-risk testing
- Agentic AI and the expansion of the attack surface from output to action
- Why red teaming and assurance are recurring rather than one-time
- A capstone scenario