AI Red Teaming & ISO 42001 Compliance Cross-Comparison

ISO 42001 Clause	How AI Red Teaming Supports Compliance	Microsoft PyRIT Capabilities	Additional Best Practices
Clause 6.1.2 & 6.1.3 – AI Risk Assessment & Treatment	Identifies AI risks through adversarial testing	Generates adversarial inputs to test model resilience	Combine with **manual testing** for broader coverage
Annex A.6.2 & A.7.4 – AI Security & Trustworthiness	Evaluates AI system robustness & data security	Tests prompt injection, data poisoning, and evasion threats	Use with **data provenance tracking** tools
Clause 6.1.4 & Annex A.5.2 – AI System Impact Assessment	Measures real-world impact of AI vulnerabilities	Assesses bias, fairness, and failure scenarios	Conduct **human-in-the-loop** impact validation
Clause 9.1 – Monitoring, Measurement & Evaluation	Ensures continuous AI model testing & validation	Runs ongoing adversarial testing & logs results	Implement **automated anomaly detection**
Clause 9.2 – Internal Audit	Provides structured audit reports on AI security	Generates logs for internal AI governance audits	Integrate with **ISO 27001 audits**
Clause 10.2 – Corrective Actions & Incident Response	Helps develop AI failure response strategies	Creates post-attack reports for mitigation planning	Build **AI-specific playbooks** for incident response

By combining AI red teaming, automated tools like PyRIT, and additional security frameworks, organizations can better meet ISO 42001 compliance requirements while strengthening AI security and governance.