4 Steps to Audit Your AI Ecosystem Before It Becomes a Liability

Step 1 : Assets

You Cannot Audit What You Haven't Mapped

What to Map

All AI and ML models in production, development, and staging

LLM integrations - internal deployments and third-party API connections

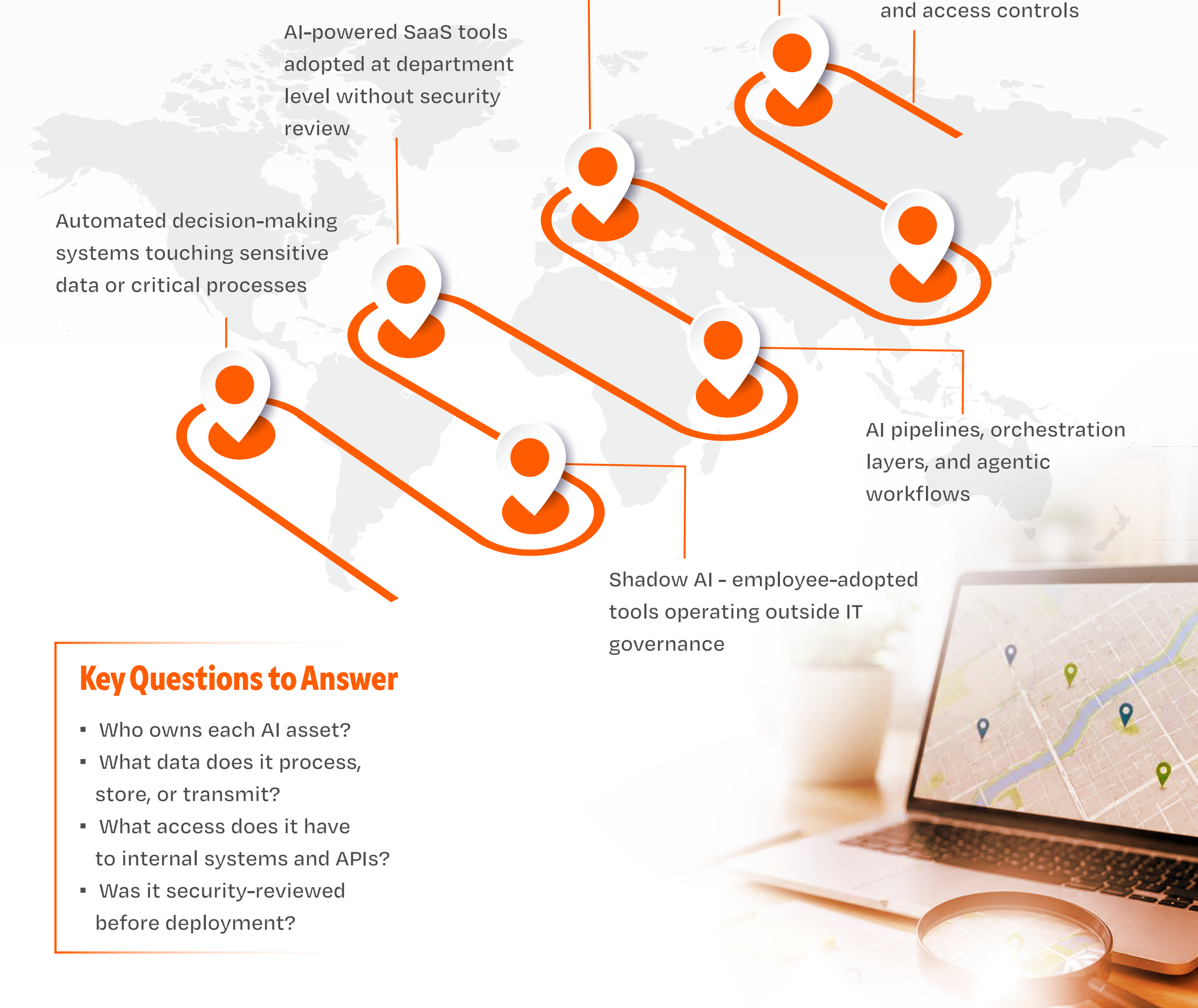
Training datasets - their source, classification, and access controls

AI-powered SaaS tools adopted at department level without security review

Automated decision-making systems touching sensitive data or critical processes

AI pipelines, orchestration layers, and agentic workflows

Shadow AI - employee-adopted tools operating outside IT governance



Key Questions to Answer

- Who owns each AI asset?
- What data does it process, store, or transmit?
- What access does it have to internal systems and APIs?
- Was it security-reviewed before deployment?

Step 2 : Threats

Map Every Vector Before an Attacker Does

What to Map

Prompt Injection
Adversarial inputs manipulating model behavior or bypassing system instructions

Data Poisoning
Corrupting training datasets to alter model behavior at inference time

Insecure API Exposure
LLM APIs without authentication, rate limits, or output filtering

Supply Chain Risk
Third-party models and open-source dependencies with unknown provenance

Indirect Prompt Injection
Malicious instructions embedded in external content processed by an AI agent

Model Inversion
Extracting sensitive training data from model outputs through repeated querying

Adversarial Examples
Crafted inputs causing misclassification in vision or NLP models

Agentic AI Misuse
Autonomous AI agents with excessive permissions executing unintended actions

Sensitive Data Leakage
Models processing PII, financial data, or IP without proper controls

MITRE ATLAS Mapping : Relevant tactics and techniques from the MITRE ATLAS framework should be mapped to each identified threat vector at this stage.

Step 3 : Controls

Test What Exists. Build What Doesn't.

Technical Controls to Test and Implement

Input and Output Security

- Input validation and sanitization against prompt injection patterns
- Output filtering to prevent sensitive data disclosure in model responses
- Content moderation layers for user-facing LLM applications

Access and Authentication

- API authentication – key management, rotation, and scope limitation
- Role-based access to model endpoints and training infrastructure
- Rate limiting and abuse detection on inference APIs

Data Security

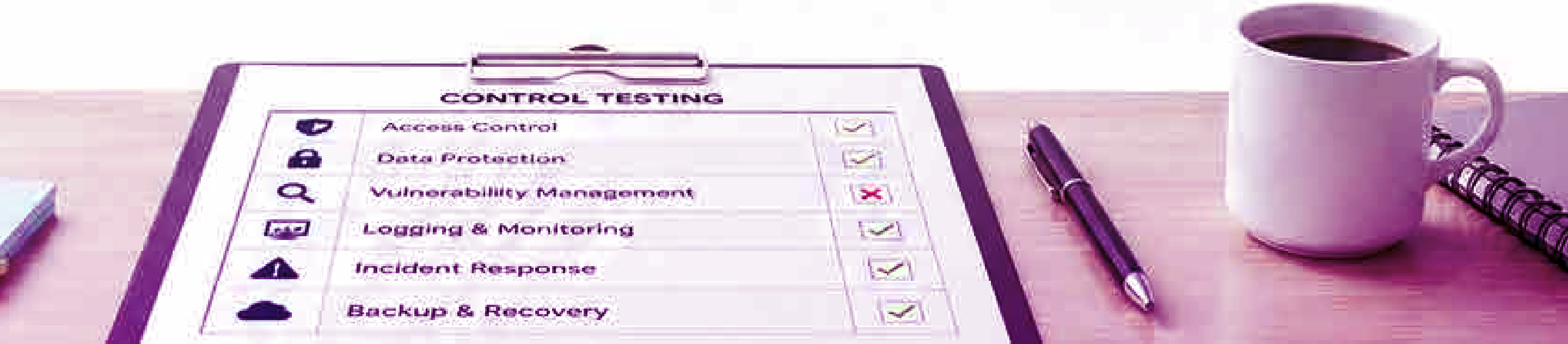
- Training data classification and access control validation
- PII detection and masking in training datasets and model outputs
- Data retention and deletion controls for AI-processed information

Model and Pipeline Security

- Secure model storage and integrity verification
- Training pipeline access controls and audit logging
- Dependency scanning for open-source model components – SBOM for AI
- Agentic AI permission scoping – least privilege for autonomous agents

Monitoring and Detection

- Anomaly detection on inference patterns – unusual query volumes or extraction behavior
- Logging of all model inputs and outputs for forensic capability
- Alerting on policy violations in real time



Step 4: Compliance

Validate Against AI Governance Rules

Frameworks to Validate Against

ISO 42001 — AI Management System

- AI risk management policy and governance structure
- Responsible AI use documentation and impact assessments
- Supplier and third-party AI risk management
- Continual improvement processes for AI security posture

NIST AI RMF

- Govern, Map, Measure, Manage framework alignment
- AI risk identification and prioritization
- Organizational accountability for AI risk decisions

DPDPA and GDPR — Data Privacy in AI Context

- Lawful basis for processing personal data in AI training and inference
- Data subject rights in automated decision -making
- Cross-border data transfer controls for AI API integrations

Industry-Specific Requirements

- RBI and IRDAI guidance on AI use in financial and insurance services
- HIPAA considerations for AI systems processing health data
- Sector-specific AI ethics and accountability obligations

EU AI Act

- Risk classification of AI systems - unacceptable, high, limited, minimal
- Conformity assessment requirements for high-risk AI deployments
- Transparency and human oversight obligations
- Technical documentation and logging requirements

The 4 Steps Are Sequential - and Continuous

The Audit Lifecycle

Step 1 - Assets →

Build and maintain an AI asset inventory. Update it with every deployment.

Step 2 - Threats →

Re-run threat modeling after AI changes, new attack methods, or security incidents.

Step 3 - Controls →

Test controls at deployment and regularly. Test AI and LLM systems more frequently.

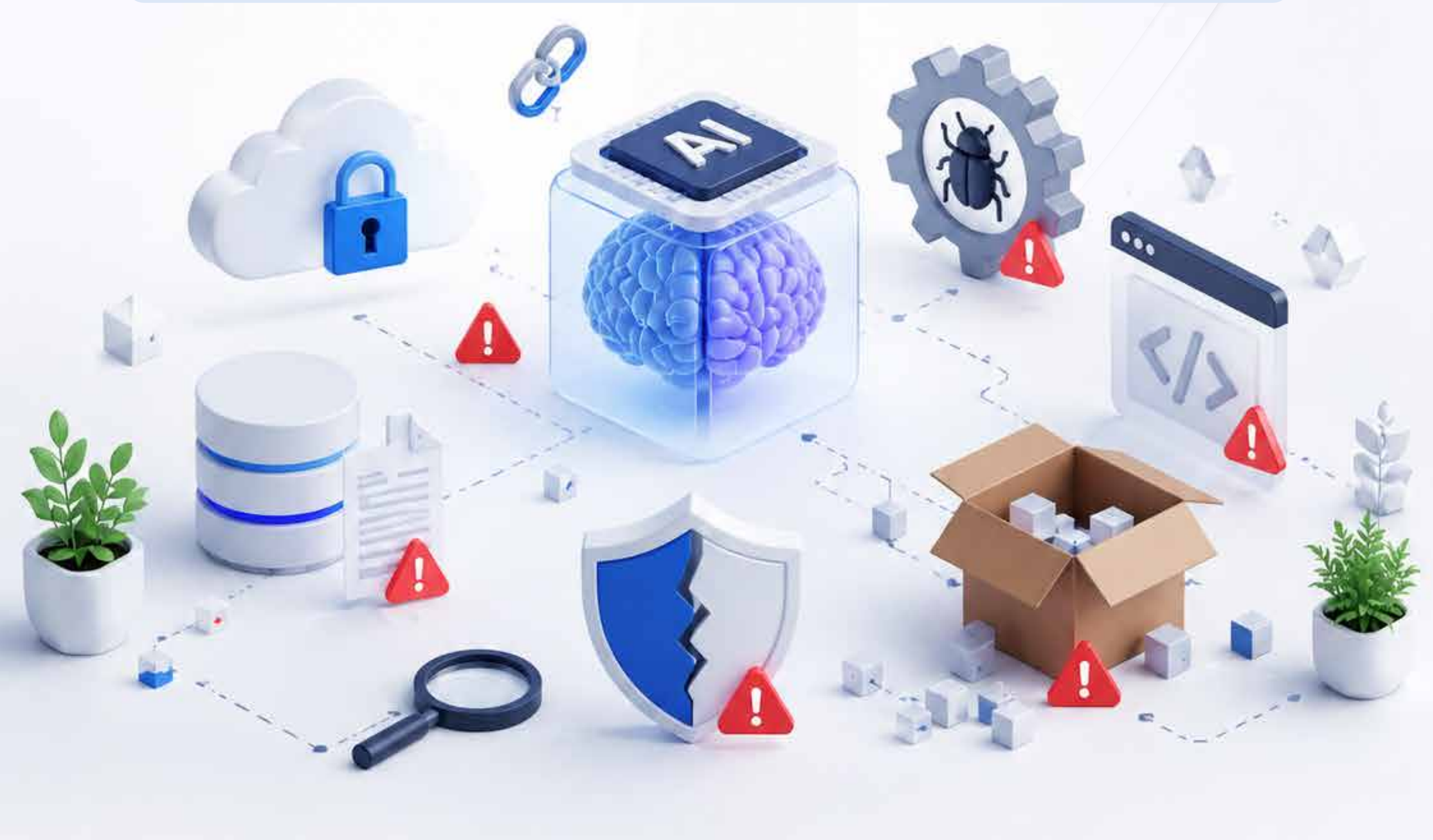
Step 4 - Compliance →

Validate compliance posture against applicable frameworks at least annually, or when regulatory guidance is updated.

What an Unaudited AI Ecosystem Actually Exposes ?

Common Gaps Found During AI Security Audits

- LLM APIs publicly accessible without authentication or rate limiting
- Employees submitting proprietary source code and financial data to external AI tools
- Agentic AI workflows with write access to production databases and email systems
- Training datasets containing unmasked PII that violates DPDPA and GDPR obligations
- Third-party AI model integrations with no vendor security assessment on record
- No logging of model inputs or outputs - zero forensic capability after an incident
- AI systems processing sensitive data with no data retention or deletion policy
- ISO 42001 gap - AI in regulated contexts with no governance controls



A Starting Point : The AI Security Audit Checklist

Assets

- Complete AI asset inventory documented and maintained
- Shadow AI discovery conducted across all business units
- Data flows for all AI systems mapped and classified

Threats

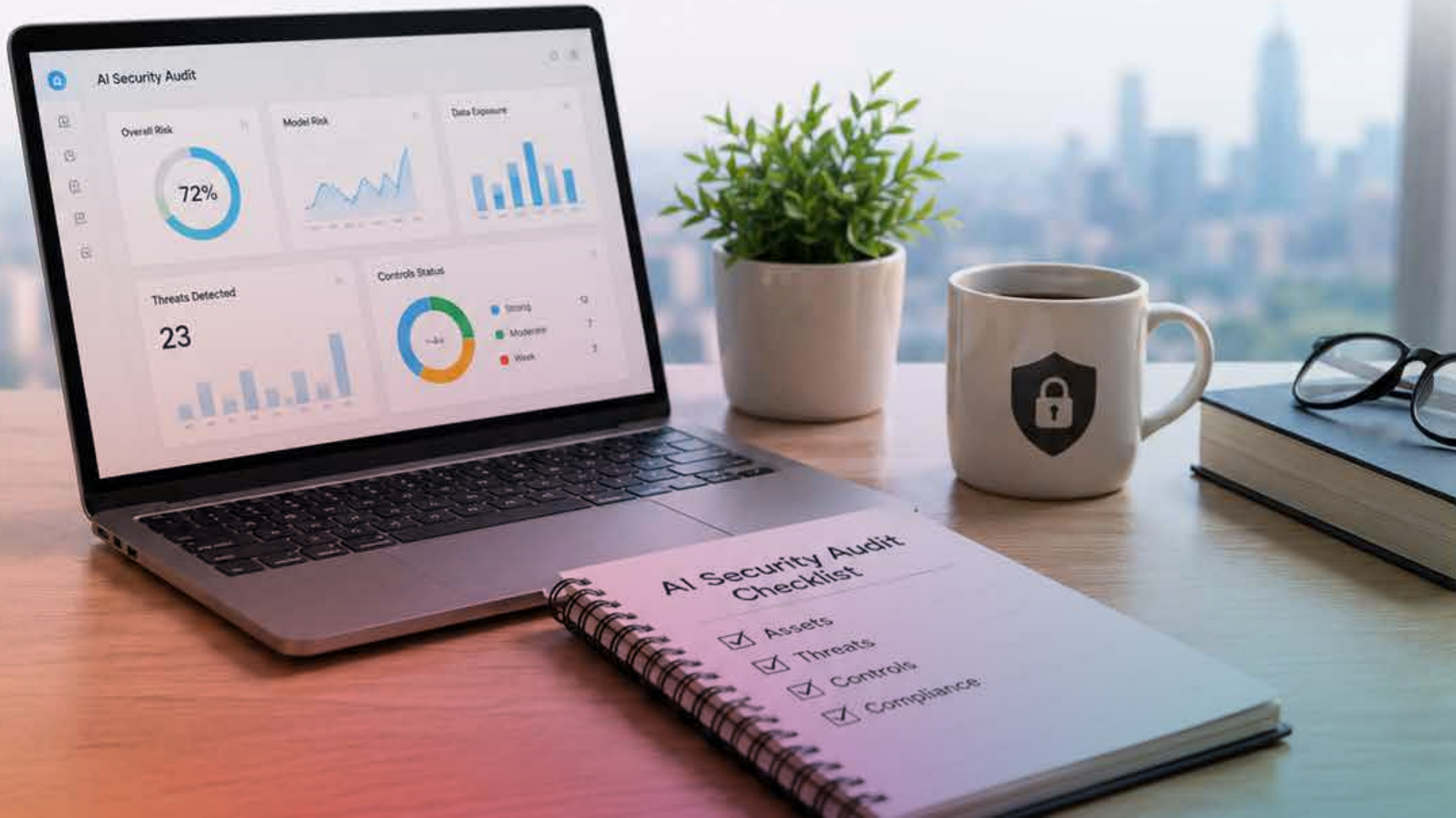
- Prompt injection testing completed for all LLM applications
- MITRE ATLAS threat model mapped to each AI asset
- Supply chain risk assessed for all third-party models and dependencies

Controls

- API authentication and rate limiting validated
- Input validation and output filtering implemented and tested
- Agentic AI permissions scoped to least privilege
- AI pipeline integrity and access controls verified
- Anomaly detection active on inference endpoints

Compliance

- ISO 42001 gap assessment completed
- EU AI Act risk classification documented for all in-scope systems
- DPDPA and GDPR obligations validated for AI data processing
- Responsible AI policy documented and approved



Has Your AI Ecosystem Been Audited Against All 4 Steps?

Our AI and LLM Security Audit covers all 4 steps

Book a Meeting