

THE AI SECURITY PLAYBOOK

20 ENTERPRISE AI THREATS AND HOW TO MITIGATE THEM

20 ENTERPRISE AI THREATS

- 01  Prompt Injection
- 02  Data Poisoning
- 03  Model Inversion
- 04  Membership Inference
- 05  Adversarial Attacks
- 06  Algorithmic Bias
- 07  Hallucinations
- 08  Sensitive Data Leakage
- 09  Insecure Plugin / Tool Use
- 10  API Abuse
- 11  Denial of Service (DoS)
- 12  Supply Chain Risks
- 13  Model Evasion
- 14  Model Extraction
- 15  Unauthorized Access
- 16  Insecure Training Pipelines
- 17  Poor Monitoring & Logging
- 18  Data Drift
- 19  Shadow AI
- 20  Unsafe Fine-Tuning

  **ALL 20 THREATS MAP TO THE 4 DISCIPLINES.**
  Strengthen the discipline.
Reduce every risk.

THE AI TRUST STACK

Security controls only hold when governance sits underneath.

1

FOUNDATION

GOVERNANCE LAYER
Defines intent, accountability, and acceptable use.

 Policies  Ownership  Accountability  Oversight

Governance sets the boundaries.

2

IDENTITY

IDENTITY LAYER
Establishes who is allowed and what they can do.

 Authentication  Authorization  RBAC / Least Privilege  Access Management

Identity grants access.

3

AOI CORE

RUNTIME VALIDATION LAYER (EXECUTION LEGITIMACY)
Validates that authority is still legitimate at the moment of execution.

 Current-State Verification  Authority Validation  State Divergence Detection  Staleness Measurement  Execution Legitimacy

AOI Integrity proves it at execution.

4

CONTROLS

SECURITY CONTROLS LAYER
Protects systems, data, and models from abuse and exploitation.

 Encryption  DLP  WAF  Monitoring  Detection  Logging

Security protects the environment.

THE 4 DISCIPLINES THAT MITIGATE ALL 20 THREATS

 **CONTROL INPUTS**

What enters the system and model.

- Input validation
- Sanitization
- Allowlists
- Data quality

 **CONTROL ACCESS**

Who can access and what they can do.

- RBAC / Least privilege
- MFA
- Authentication
- Permissions

 **OBSERVE RUNTIME BEHAVIOR**

Watch what actually happens in production.

- Monitoring
- Logging
- Drift detection
- Red teaming
- Anomaly detection

 **KEEP HUMANS ACCOUNTABLE**

Ensure human oversight and decision integrity.

- Human review
- Confidence scoring
- Access reviews
- Escalation paths



SHADOW AI
THE ONLY RISK THAT IS PURE GOVERNANCE

Employees using unapproved AI tools cannot be solved by security controls alone. It requires governance, awareness, and accountability.

19 OF THESE THREATS HAVE TECHNICAL FIXES.
THIS ONE DOESN'T.

KEY TAKEAWAY

Twenty named threats.
Four enduring disciplines.
One truth: authority is assumed.

AOI Integrity proves it at execution.



AOI

ARCHITECTURE OF INTEGRITY

AUTHORITY IS ASSUMED.

AOI INTEGRITY PROVES IT AT EXECUTION.



Not just identity.
Not just security.



Runtime Governance.
Execution Legitimacy.



Proving authority when it matters most.