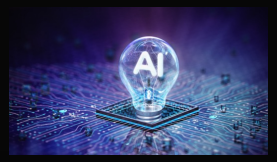
TrendAI™
AI Fearlessly



Defining AI Stack Telemetry

Detection, Incident Response, Compliance, Governance

Global Digital Foundation, March 4th 2026, Barcelona, Spain

David Girard, Sr. Director Global
AI Security, AI Alliances, AI Factory

CoSAI Project Governing Board

The AI Observability Crisis

Case Study

MS 365 Copilot Agent Exploit - EchoLeak

Incident
Researchers demonstrated a prompt injection attack via email

Impact
Sensitive data exfiltration

Missing Observability
No user-facing audit trail, no visibility into data access, opaque tool execution, limited security telemetry, no decision explainability

Outcome
Demonstrated the need for security visibility onto agent behavior

Why Now: The Perfect Storm

Rapid AI Adoption
Enterprise AI Deployments accelerating

Regulatory Pressure
EU AI Act, emerging compliance requirements

Visibility Gap
Traditional monitoring blind to AI behavior

Security Incidents
Growing Attack Surface, novel threats

Why Now? The Convergence of Pressures

Four forces demanding comprehensive AI telemetry

Regulatory Mandates

- EU AI ACT
- NIST AI RMF
- ISO 42001

Security Threats

- MITRE ATLAS
  - 66+ techniques
  - 14 new agentic
- OWASP
  - Top 10 LLM
  - Top Agentic

Production Scale

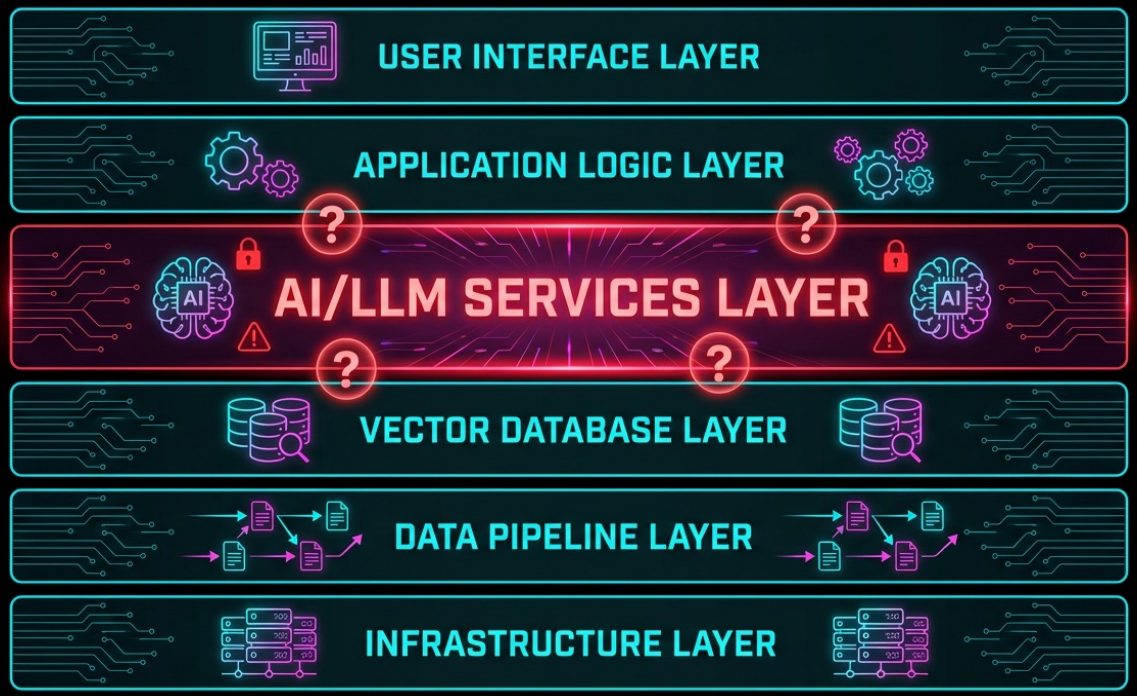
- Experiments → Mission critical
- Customer facing
- Revenue impacting

Agentic Future

- **Autonomous agents**
- **Tool calling**
- **Multi-agents systems**
- **Agentic Identity proliferation**

The AI Stack Visibility Gap

Traditional monitoring tools miss critical AI specific layers



Vector DB like FAISS has no logging at all.

Some AI Services provide basic logs with low visibility into tool access or reasoning.

Real-World Failures & Organizational Barriers

High-Profile AI Failures

SAMSUNG

Incident:
Engineers leaked proprietary code via ChatGPT

Gap:
No data exfiltration monitoring

Air Canada

Incident:
Chatbot made unauthorized promises

Gap:
No conversation logging

Chevrolet

Incident:
Dealer chatbot agreed to sell car for $1

Gap:
No output validation telemetry

5 Barriers to AI Logging

1. **Privacy Concerns**
   Fear of logging sensitive user data
2. **Performance Impact**
   Perceived latency from logging
3. **Cost Considerations**
   Storage and processing overhead
4. **Technical Complexity**
   No standardized AI telemetry format
5. **Organizational Silos**
   ML teams disconnected from SecOps

Framework Landscape Overview

Analyzing AI security frameworks for telemetry requirements

Framework	Focus Area	Telemetry Depth	Maturity
OWASP LLM Top 10	Application vulnerabilities	**Medium**	Established
OWASP Agentic Security	Agent-specific risks	**High**	Emerging
MITRE ATLAS	Adversarial ML tactics	**High**	Established
CSA AI Controls Matrix	Cloud AI governance	**Medium**	Established
NIST AI RMF	Risk management	**Low**	Established
ISO 42001	AI management systems	**Low**	New
OpenTelemetry GenAI SIG	Gen AI Telemetry	**Low**	Low

OWASP LLM & Agentic + MITRE ATLAS

OWASP Top 10 LLM & Agentic

Key Risks:

Top 10 LLM

- Prompt Injection
- Sensitive Information Disclosure
- Excessive Agency
- Supply Chain Vulnerabilities

Top 10 Agentic

- Excessive Agency
- Tool Misuse
- Unsafe Delagation

Telemetry Requirements:

- Input/output logging for injection detection
- Data flow monitoring for PII exposure
- Agent action audit trails
- Dependency and model provenance tracking

Coverage: Application & Agent Layer

MITRE ATLAS

Key Tactics:

- Reconnaissance
- Initial Access
- ML Attack Staging
- Exfiltration

Telemetry Requirements:

- Model query patterns and rate limiting
- Training data access logs
- Model behavior drift detection
- Adversarial input signatures

Coverage: Adversarial ML Lifecycle

ISO 42001 + Cross-Framework Telemetry Mapping

ISO 42001
AI Management Systems

Key Clauses:

- 6.1: Risk Assessment
- 7.5: Documented Information
- 8.4: AI System Development
- 9.1: Monitoring & Measurement

Telemetry Implication:

Requires evidence of AI system monitoring for certification

Telemetry Category	OWASP	MITRE	CSA	NIST	ISO
Input/Output Logging	🟢	🟢	🟢	🔴	🟢
Model Performance	🔴	🟢	🟢	🟢	🟢
Data Lineage	🟢	🔴	🟢	🟢	🟢
Security Events	🟢	🟢	🟢	🔴	🔴
Agent Actions	🟢	🟢	🔴	🔴	🔴

Threat & Governance Mapping

Threat Detection

Interaction	Prompt injection, social engineering
Model	Model theft attempts, evasion attacks
Security	All OWASP LLM Top 10 threats
Data	Data exfiltration, poisoning
Operational	DoS, resource abuse

Compliance & Governance

Interaction	Audit trails, user consent evidence
Model	Performance accountability, SLAs
Security	Incident response, breach notification
Data	GDPR, data sovereignty proof
Operational	Cost governance, capacity planning

Proposed AI Telemetry Standard

Five essential categories with implementation priorities

HIGH

1

Interaction Telemetry

- User prompts
- System responses
- Session context
- Conversation chains

HIGH

2

Model Telemetry

- Token counts
- Latency metrics
- Confidence scores
- Model version

HIGH

3

Security Telemetry

- Injection attempts
- Guardrail triggers
- Access violations
- Anomaly flags
- Exfiltration Indicator

MEDIUM

4

Data Telemetry

- RAG retrievals
- Source citations
- PII detection
- Data lineage

MEDIUM

5

Operational Telemetry

- Resource usage
- Error rates
- Queue depths
- Cost metrics

Introducing AITF

AI Telemetry Framework — Bridging Observability and Security

OpenTelemetry GenAI

Traces, Metrics & Events
gen_ai.* semantic conventions
Model, Agent, Tool spans

AITF

OCSF Security Schema

Normalized Security Events
New Category 7: AI/ML Activity
SIEM/XDR native ingestion

AITF Architecture

L1 Instrumentation	OTel GenAI SDK + AITF Extensions
L2 Collection	OTel Collector + Security Processor
L3 Normalization	AITF OCSF Mapper
L4 Analytics	XDR / SIEM / Dashboards

One pipeline. Two standards. Full coverage from developer to SOC.

AITF Event Classes

OCSF Category 7: AI/ML Activity — Eight classes covering the full AI attack surface

7001
AI Inference

Prompts, responses, token usage, model params

7002
AI Agent Activity

Orchestration, planning, delegation, boundaries

7003
AI Tool Execution

MCP skills, functions, API calls, permissions

7004
AI Data Retrieval

RAG pipeline, vector DB, knowledge base access

7005
AI Security Finding

Guardrails, content filters, policy engine results

7006 v1.1
AI Supply Chain

Model provenance, skill integrity, AI-BOM, signing

7007 v1.1
AI Integrity

Config drift, agent def tampering, guardrail bypass

7008 v1.1
AI Identity

Agent auth, credentials, delegation chains, trust

Natively ingestible by OCSF-compatible SIEMs: TrendAI Vision One | Splunk | AWS Security Lake | Datadog

TrendAI

What Makes AITF Different

Beyond inference monitoring — securing the entire AI lifecycle

AI Supply Chain Security

- Model hash verification
- Skill provenance tracking
- AI-BOM generation
- Registry trust scoring
- Dependency monitoring
- Signature verification

Integrity Monitoring

- Agent definition baselines
- Skill manifest mutation detect
- Guardrail policy drift alerts
- Config change approval flow
- Prompt template tracking
- Automated rollback support

Agentic Identity

- Agent authentication (mTLS)
- Delegation chain tracking
- Credential lifecycle mgmt
- Zero-trust for AI agents
- A2A identity federation
- Privilege escalation alerts

AITF doesn't just monitor inference — it secures the entire AI supply chain and identity layer.

AITF Implementation Roadmap

Quick Wins
0-30 Days

- ✓ Enable basic LLM API logging
- ✓ Inventory all AI touchpoints
- ✓ Define retention policies
- ✓ Establish baseline metrics

OUTCOME:
Immediate visibility into AI usage

Medium-Term
1-6 Months

- ✓ Implement security telemetry
- ✓ Integrate with SIEM/XDR
- ✓ Deploy anomaly detection
- ✓ Build compliance dashboards

OUTCOME:
Proactive threat detection capability

Long-Term
6-18 Months

- ✓ Full telemetry standard adoption
- ✓ Automated response playbooks
- ✓ Continuous compliance monitoring
- ✓ AI-specific SOC capabilities

OUTCOME:
Mature AI security operations

"You can't secure what you can't see"

"AITF: Making AI visible to the defenders who need it most"

Security Axiom for the AI Era

AI telemetry is not optional—it's the foundation of AI security and governance.

Start Today

- Audit your AI stack visibility
- Prioritize high-value telemetry
- Build toward the standard

Questions?

Thank You

TrendAI™
AI Fearlessly

David Girard

David_Girard@trendmicro.com