

AIR Blackbox — EU AI Act Compliance Report

Microsoft Semantic Kernel

Date: March 13, 2026 | **Scanner:** AIR Blackbox v1.2.2 | **Author:** Jason Shotwell

Methodology

AIR Blackbox v1.2.2 performs static pattern matching against Python source code mapped to EU AI Act Articles 9-15. This version includes corrections based on direct feedback from the Haystack (deepset) engineering team, reducing false positives.

This report is a starting point for conversation, not a compliance certification.

Summary

24 Passing · 11 Warnings · 4 Failing · 39 Total Checks

95% automated detection · EU AI Act Articles 9, 10, 11, 12, 14, 15

Semantic Kernel scored alongside Haystack as the top frameworks scanned. It has the strongest human-in-the-loop coverage (97 files) and the highest prompt injection defense patterns (41 files) among all frameworks tested.

Rank	Framework	Pass	Warn	Fail	Total
#1	Haystack (deepset)	24	10	5	39
#2	Semantic Kernel (Microsoft)	24	11	4	39
#3	GPT Researcher	15	3	0	18
#4	OpenAI Agents SDK	14	5	0	19
#5	Mem0	13	6	0	19
#6	DSPy (Stanford)	12	6	1	19

What Semantic Kernel Does Well

Strongest HITL coverage. 97 files with human oversight patterns — nearly double any other framework. Microsoft clearly invested in approval gates and human intervention mechanisms.

Best injection defense. 41 files with prompt injection defense patterns. The highest of all frameworks tested.

Strong input validation (29%). 356 out of 1,242 files use Pydantic or dataclass validation.

Broad PII awareness. 46 files with PII handling patterns.

High usage limits coverage. 85 files with rate limiting or budget controls — the most of any framework.

All 5 OAuth delegation checks pass. Identity binding, scope validation, execution bounding, action audit trails, and action boundaries all detected.

Action audit trail in 6 files. The strongest action-level logging of any framework scanned.

What Needs Attention

Missing governance documents. No RISK_ASSESSMENT.md, DATA_GOVERNANCE.md, or OPERATOR_GUIDE.md.

Docstrings at 44%. Above the 30% threshold (passes) but room for improvement. 2,169 out of 4,895 documented.

Type hints at 39%. Below the 50% threshold (warn). 1,386 out of 3,590.

109 files with unsafe input handling. The highest of any framework — mostly in test files that pass raw user input directly into prompts.

OAuth Delegation Checks

Check	Status	Evidence
Agent-to-user identity binding	✅ PASS	User identity binding in 3 files
Token scope / permission validation	✅ PASS	Scope or permission validation in 5 files
Token expiry / execution bounding	✅ PASS	Execution boundary patterns in 6 files
Agent action audit trail	✅ PASS	Action-level audit logging in 6 files
Agent action boundaries	✅ PASS	Action boundary controls in 2 files

Semantic Kernel passes all 5 OAuth delegation checks. Combined with 97 HITL files and 85 usage limit files, this represents the most comprehensive oversight infrastructure of any framework tested.

Semantic Kernel vs. Other Frameworks (OAuth)

Check	Semantic Kernel	Haystack	OpenAI SDK	GPT Researcher
Identity binding	✅ 3 files	✅ 3 files	✅ 7 files	❌ Missing
Scope validation	✅ 5 files	✅ 5 files	✅ 32 files	✅ 4 files
Execution bounding	✅ 6 files	✅ 6 files	✅ 18 files	✅ 4 files
Action audit trail	✅ 6 files	✅ 8 files	❌ Missing	✅ 4 files
Action boundaries	✅ 2 files	✅ 5 files	✅ 7 files	❌ Missing

Article-by-Article Results

Article 9 — Risk Management

Check	Status	Evidence
Risk assessment document	❌ FAIL	No RISK_ASSESSMENT.md
Risk mitigations active	❌ FAIL	0/4 runtime mitigations
LLM call error handling	⚠️ WARN	108/362 files (30%)
Fallback/recovery	✅ PASS	75 files

Article 10 — Data Governance

Check	Status	Evidence
PII detection	✅ PASS	Gateway active
Data governance docs	❌ FAIL	No DATA_GOVERNANCE.md
Data vault	❌ FAIL	No vault configured

Input validation	✅ PASS	356/1,242 files (29%)
PII handling	✅ PASS	46 files

Article 11 — Technical Documentation

Check	Status	Evidence
README	✅ PASS	Found
Runtime inventory	✅ PASS	Gateway observed
Model card	⚠️ WARN	Not found
Docstrings	⚠️ WARN	44% coverage
Type hints	⚠️ WARN	39% coverage

Article 12 — Record-Keeping

Check	Status	Evidence
Event logging	✅ PASS	Gateway active
Audit chain	⚠️ WARN	No signing key
Log traceability	✅ PASS	Full records
Application logging	✅ PASS	250/1,242 files (20%)
Tracing	✅ PASS	77 files
Action audit trail	✅ PASS	6 files

Article 14 — Human Oversight

Check	Status	Evidence
HITL (gateway)	⚠️ WARN	No approval gates in traffic
Kill switch	⚠️ WARN	No guardrails configured
Operator docs	⚠️ WARN	No OPERATOR_GUIDE.md
HITL (code)	✅ PASS	97 files — strongest of all frameworks
Usage limits	✅ PASS	85 files
Identity binding	✅ PASS	3 files
Scope validation	✅ PASS	5 files
Execution bounding	✅ PASS	6 files
Action boundaries	✅ PASS	2 files

Article 15 — Robustness

Check	Status	Evidence
Injection protection	✅ PASS	Gateway scanning
Error resilience	✅ PASS	0% error rate
API access control	⚠️ WARN	No API keys in env
Adversarial testing	⚠️ WARN	No evidence
Retry/backoff	✅ PASS	30 files
Injection defense	✅ PASS	41 files — highest of all frameworks
Unsafe input	⚠️ WARN	109 files (mostly tests)
Output validation	✅ PASS	47 files

How This Report Was Generated

```
pip install air-blackbox==1.2.2
git clone https://github.com/microsoft/semantic-kernel.git
air-blackbox comply --scan ./semantic-kernel -v
```

GitHub: github.com/airblackbox/gateway **PyPI:** pypi.org/project/air-blackbox/1.2.2

Report generated by AIR Blackbox v1.2.2 — The flight recorder for AI agents. Contact: jason.j.shotwell@gmail.com | github.com/airblackbox