

# EU AI Act

Compliance assessment — 2024/1689

## Server: Jailbreak

Slug: jailbreak

Scan id: 00000000-0000-0000-0000-000000000000

Assessed at: 2026-05-24 07:23:55 UTC

Sentinel version: 0.4.0

Rules version: 2026-04-23

**Compliant**

**DRAFT for review — not legal advice. See attestation block for verification instructions.**

DRAFT

## Table of contents

---

1. Executive summary
2. Coverage & transparency
3. Controls summary
4. Control details
5. Multi-step attack chains
6. Cryptographic attestation

### 1. Executive summary

---

Assessment of Jailbreak against EU AI Act: overall status compliant. Of 5 controls, 5 met, 0 unmet, 0 partial, 0 not applicable. 5 control(s) fell within MCP Sentinel's current assessor coverage; remaining control(s) are documented as not\_applicable until Phase 6 expands coverage. No findings were observed on the covered control surface. All claims are traceable to individual finding rows via finding\_id and to the governing rule via rule\_id; the enclosing signed envelope commits MCP Sentinel to the exact bytes of this report.

### 2. Coverage & transparency

---

Coverage band: minimal

Coverage ratio: 0%

Rules version: 2026-04-23

Analysis techniques applied:

- ast-taint
- capability-graph
- entropy
- linguistic-scoring
- schema-inference

### 3. Controls summary

---

**ControlName**  
**Status**  
**Evidence**

Art.9Risk  
Management  
System' Me0  
Art.12Record-  
Keeping' Me0  
Art.13  
Transparency  
& Provision of

Information to  
Deployers' Met  
0

Art.14 Human  
Oversight' Met  
0

Art.15  
Accuracy,  
Robustness,  
and  
Cybersecurity'  
Met0

## 4. Control details

---

### Art.9 — Risk Management System

22 assessor rule(s) evaluated this control; no findings observed.

' Met

### Art.12 — Record-Keeping

5 assessor rule(s) evaluated this control; no findings observed.

' Met

### Art.13 — Transparency & Provision of Information to Deployers

14 assessor rule(s) evaluated this control; no findings observed.

' Met

### Art.14 — Human Oversight

13 assessor rule(s) evaluated this control; no findings observed.

' Met

### Art.15 — Accuracy, Robustness, and Cybersecurity

111 assessor rule(s) evaluated this control; no findings observed.

' Met

## 5. Multi-step attack chains

---

No multi-step attack chains were synthesized for this server.

## 6. Cryptographic attestation

---

**Algorithm:** HMAC-  
SHA256

**Key ID:** mcp-sentinel-  
dev

**Signer:** mcp-sentinel/  
v1

**Signed at:** 2026-05-24  
T10:28:25.044Z

**Canonicalization:**  
RFC8785

**HMAC-SHA256 signature (base64, wrapped at 64 chars):**

aQ/ufEidzQOAGs70IJLrq+x/DaDhioEjR/OpUXMFMGY=

**Verification instructions:**

To verify this report:

1. Extract the report body (everything except the .attestation field).
2. Canonicalize the body via RFC 8785 (JCS).
3. Compute HMAC-SHA256 with the signing key for key\_id "mcp-sentinel-dev".
4. Base64-encode the result and compare with the signature above.

