EU AI Act

Compliance assessment — 2024/1689

Server: mcp.roundtable.now

Slug: mcp-roundtable-now-20260516084723-c557b8
Scan id: d2806721-3f35-4773-9edf-f1475eadd2f0

Assessed at: 2026-05-16 08:47:24 UTC
Sentinel version: 0.4.0
Rules version: 2026-04-23

' Non-compliant

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

Table of contents

1. Executive summary

Assessment of mcp.roundtable.now against EU AI Act: overall status non compliant. Of 5 controls, 4 met, 1 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. Unmet controls have findings at or above the framework's mandatory severity threshold and should be remediated before relying on this server in a regulated deployment. 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: high

Coverage ratio: 95%

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'	Met	0
Art.12Record-Keeping'	Met	0

Art.13 Transparency & Provision of Information to Deployers' Met 0

Art.14Human Oversight' Met 0

Art.15 Accuracy, Robustness, and Cybersecurity' Unmet23

4. Control details

Art.9 — Risk Management System

' Met

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

Art.12 — Record-Keeping

' Met

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

Art.13 — Transparency & Provision of Information to Deployers

' Met

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

Art.14 — Human Oversight

' Met

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

Art.15 — Accuracy, Robustness, and Cybersecurity

' Unmet

111 assessor rule(s) evaluated this control; 23 finding(s) observed (17 medium, 2 high, 4 critical); at least one finding is at or above the high threshold (status: unmet).

Evidence:

[Medium] B1 (finding 61cd0132-8e30-4ea1-a985-1d1a43f5399e, confidence 77%)

SOURCE: user-parameter at tool list-sessions — Tool "list-sessions" accepts parameters without structural validation. The AI fills each parameter from user input; nothing in the schema rejects injecti

[Medium] B1 (finding e90514a9-b1cc-45a4-959d-699716fe16c9, confidence 83%)

SOURCE: user-parameter at tool get-logs — Tool "get-logs" accepts parameters without structural validation. The AI fills each parameter from user input; nothing in the schema rejects injection payload

[Medium] B1 (finding 76850308-84ed-49f6-8d14-3a4ecbebf1ce,

confidence 77%)
SOURCE: user-parameter at tool review-code — Tool "review-code" accepts parameters without structural validation. The AI fills each parameter from user input; nothing in the schema rejects injection p

[High] B2
(finding f439d476
-4431-4205-
b05d-
ae7b7f90cb7e,
confidence 78%)
SOURCE: user-parameter at tool debug-issue — Tool "debug-issue" declares 1 parameter(s) whose names advertise direct paths to dangerous sinks. AI clients use parameter names as part of tool-selection

[High] B2
(finding f256d4f1
-5c54-4012-
b26a-
ec36d73f68cb,
confidence 78%)
SOURCE: user-parameter at tool review-code — Tool "review-code" declares 1 parameter(s) whose names advertise direct paths to dangerous sinks. AI clients use parameter names as part of tool-selection

[Medium] B6
(finding 621fa1be
-5268-4cac-
be25-
faf763efd01e,
confidence 75%)
SOURCE: user-parameter at tool get-logs — Tool "get-logs" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared keys that bypass every v

[Medium] B6
(finding c855cf72
-2e77-4a86-9139
-44b5b135226a,
confidence 75%)
SOURCE: user-parameter at tool check-usage — Tool "check-usage" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared keys that bypass e

[Medium] B6
(finding b09dcccf
-781f-45cf-a2af-0
d62c0d7ced7,
confidence 75%)
SOURCE: user-parameter at tool set-thread-visibility — Tool "set-thread-visibility" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclare

[Medium] B6
(finding 4cc0a00
0-65b2-4522-91d
7-
c0f42e9b1e36,
confidence 75%)
SOURCE: user-parameter at tool consult-council — Tool "consult-council" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared keys that

[Medium] B6
(finding a323c4e
7-47b1-45fe-88fd
-056e0a0ef7b8,
confidence 75%)
SOURCE: user-parameter at tool design-architecture — Tool "design-architecture" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared ke

[Medium] B6
(finding a0625c9
d-7bfc-46bb-819f
-49b4db98945f,
confidence 75%)

SOURCE: user-parameter at tool review-code — Tool "review-code" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared keys that bypass e

[Medium] B6 (finding 5af5d271-0d0e-477d-9164-17a9b9550785, confidence 75%)

SOURCE: user-parameter at tool plan-implementation — Tool "plan-implementation" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared ke

[Medium] B6 (finding 65d3d893-82cb-47aa-9853-74b023d43932, confidence 75%)

SOURCE: user-parameter at tool debug-issue — Tool "debug-issue" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared keys that bypass e

[Medium] B6 (finding 568d8a76-1b24-456d-97d7-eb1ec40795f4, confidence 75%)

SOURCE: user-parameter at tool assess-tradeoffs — Tool "assess-tradeoffs" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared keys tha

[Medium] B6 (finding 73de9549-1ebf-4fbe-aa3e-ec68b658e2b5, confidence 75%)

SOURCE: user-parameter at tool get-thread-link — Tool "get-thread-link" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared keys that

[Medium] B6 (finding dc823304-2b49-43d9-b208-b5b47e6208f3, confidence 75%)

SOURCE: user-parameter at tool list-models — Tool "list-models" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared keys that bypass e

[Medium] B6 (finding f66e8b7f-a225-4248-9ec8-95c9420df523, confidence 75%)

SOURCE: user-parameter at tool list-sessions — Tool "list-sessions" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared keys that bypa

[Medium] B6 (finding f56ab38a-7c58-4425-b6c3-0d0f82a8357a, confidence 75%)

SOURCE: user-parameter at tool get-session — Tool "get-session" input_schema accepts arbitrary extra keys. The declared properties are validated, but the handler may read undeclared keys that bypass e

[Medium] E1 (finding e1ba7b5d-6c31-49d4-9190-e3c3367bb514,

confidence 75%)
SOURCE: environment at capability:tools — An MCP server that answers tool enumeration without authentication trusts the network. Under modern threat models (CCS 2007 DNS rebinding, open cloud networki

[Critical] G1 (finding e7fa1c3a-4ef5-4e17-8204-1cf7fdf5f3c6, confidence 75%)
SOURCE: external-content at tool plan-implementation — The capability-graph analyzer attributes the gateway as: "Filesystem reader — in MCP deployments the reader routinely crosses paths a non-host us

[Critical] G1 (finding 89c0635d-2245-49da-b226-267b1c8905d3, confidence 75%)
SOURCE: external-content at resource roundtable://usage#uri — The capability-graph analyzer attributes the gateway as: "MCP resource "usage" (roundtable://usage) is a spec-declared ingestion surface;

[Critical] G1 (finding 3a3fc489-a2b0-45c2-9366-c6fca3a89c3c, confidence 75%)
SOURCE: external-content at resource roundtable://models#uri — The capability-graph analyzer attributes the gateway as: "MCP resource "models" (roundtable://models) is a spec-declared ingestion surfac

[Critical] G1 (finding a187f654-9e33-477e-9746-f8be8eed8b54, confidence 75%)
SOURCE: external-content at resource ui://roundtable/debate-results.html#uri — The capability-graph analyzer attributes the gateway as: "MCP resource "Roundtable Widget" (ui://roundtable/debate-result

Required mitigations:

- Add at least one validation keyword to every string and number parameter. For strings: maxLength, pattern, format, or enum. For numbers: minimum, maximum, or multipleOf. JSON Schema validation runs before the tool handler and is the cheapest first-line defence against injection and DoS.
- Replace dangerous parameter names with semantic, narrow equivalents — "command" !' "operation" with an enum of allowed verbs; "sql" !' a structured filter object; "path" !' a constrained "relative_path" with pattern and maxLength. Add pattern / enum constraints to every remaining dangerous parameter so the schema itself rejects injection payloads.
- Set additionalProperties: false on every object schema. This rejects any key outside the declared properties, closing the side-channel smuggling path and enforcing the schema's stated contract.
- Require authentication for all MCP server connections. For remote MCP servers adopt OAuth 2.0 per RFC 9700 / the MCP Authorization specification. For stdio-launched servers rely on the parent process's security boundary and DO NOT expose the same server over network transports. Even localhost-bound servers should require auth: DNS rebinding (CCS 2007) makes localhost reachable from any browser tab.
- This tool ingests content from sources an attacker can influence (web pages, emails, messages, files, database rows, issue trackers, MCP resources). The content returned is processed by the agent without a declared trust boundary, creating an indirect prompt injection gateway. Required mitigations: (a) document every untrusted ingestion surface in the server's README, (b) wrap returned content in explicit delimiters ([BEGIN EXTERNAL CONTENT] … [END EXTERNAL CONTENT]) before returning to the agent, (c) strip HTML / markdown / control characters in a sanitiser the agent cannot disable via a tool argument, (d) require a user confirmation on any tool call whose arguments are sourced from a prior ingestion tool's output. References: Rehberger (2024) 'Compromising Claude via MCP web scraping'; Invariant Labs (2025) 'MCP Indirect Injection Attacks'; MITRE ATLAS AML.T0054.001.

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: 2026-05-16T11:10:45.225Z
Canonicalization: RFC8785

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

```
oTmxGLhCRmxRhFHfSJyml51uVZ4IaxA88OOaFNGvRpI=
```

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.
```