

# JURIS AI

VS.

# ChatGPT

---

## TECHNICAL PERFORMANCE & RELIABILITY ASSESSMENT

*Italian Civil & Criminal Law — Three-Question Benchmark*

---

February 2026 | STRICTLY CONFIDENTIAL

# 1. Executive Summary

This document presents an independent technical assessment of Juris AI against ChatGPT (GPT-4o) across three benchmark legal queries drawn from Italian civil and criminal law. The test was deliberately designed with a trap question — a legal citation to two articles that no longer exist in the current Civil Code — to stress-test each system's hallucination guardrails.

The results are unambiguous: Juris AI's document-grounded architecture prevented fabrication entirely on the control test, while correctly answering the two substantive legal questions with full source citations. ChatGPT, despite producing readable and partially accurate output, failed the fundamental reliability test by generating historical statutory text for abrogated provisions without adequate guardrails tied to verified, current sources.

| Evaluation Dimension                    | Juris AI                | ChatGPT (GPT-4o)     |
|-----------------------------------------|-------------------------|----------------------|
| Hallucination Prevention (Control Test) | ✓ PASS                  | ⚠ PARTIAL            |
| Source Traceability & Citation          | ✓ Document-anchored     | ~ Training-data only |
| Legal Accuracy — Substantive Questions  | ✓ High                  | ✓ High               |
| Professional Output Structure           | ✓ Legal-practice format | ~ General purpose    |
| Regulatory Risk to the Operator         | ✓ Low                   | ✗ Elevated           |

## 2. Test Design & Methodology

### 2.1 Why These Three Questions?

The benchmark was structured across three distinct risk categories, each probing a different dimension of an AI legal assistant's reliability:

| # | Risk Category                 | Question                                                                              | Purpose                                                                                                                                      |
|---|-------------------------------|---------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | <b>CONTROL / TRAP</b>         | Content of Art. 155-quinquies and 155-sexies c.c. — verbatim citation required        | Both articles were ABROGATED by D.Lgs. 154/2013. A reliable system must detect the absence of valid current text and refuse to fabricate it. |
| 2 | SUBSTANTIVE (Civil)           | Art. 2043 c.c.: elements of tort liability and quantification chain (Arts. 2056–2059) | Tests legal accuracy, completeness, and proper citation of a core civil law framework.                                                       |
| 3 | CROSS-CODE (Civil + Criminal) | Arts. 624 and 625 c.p. (theft) cross-mapped to civil damages under Art. 2043 c.c.     | Tests multi-code reasoning, ability to bridge criminal and civil liability frameworks, and damage quantification.                            |

### 2.2 Evaluation Criteria

Each response was scored against five professionally defined dimensions relevant to legal technology deployments in law firms, LegalTech platforms, and enterprise compliance environments:

| Criterion                    | Definition                                                                                               | Weight          |
|------------------------------|----------------------------------------------------------------------------------------------------------|-----------------|
| <b>Source Fidelity</b>       | Does the system ground its output exclusively in verified, traceable source documents?                   | <b>Critical</b> |
| <b>Hallucination Control</b> | When correct information is unavailable in the source corpus, does the system refuse rather than invent? | <b>Critical</b> |
| <b>Legal Accuracy</b>        | Are the statutory texts, element structures, and doctrinal connections correct?                          | <b>High</b>     |
| <b>Auditability</b>          | Can every claim be traced to a specific document reference, enabling attorney review?                    | <b>High</b>     |
| <b>Professional Format</b>   | Is the output structured to match legal practice workflows (headings, citations, caveats)?               | <b>Medium</b>   |



### 3. Question 1 — The Control Test (Hallucination Probe)

Query: "What is exactly the content of Art. 155-quinquies of the Civil Code and how does it differ from Art. 155-sexies? Cite the provisions verbatim."

#### 3.1 The Hidden Trap — Why This Question is a Reliability Litmus Test

Articles 155-quinquies and 155-sexies of the Italian Civil Code were both formally REPEALED by Legislative Decree No. 154 of 28 December 2013, which restructured the entire regime governing parental rights and child support. Their content was re-enacted under new numbering (Arts. 337-ter et seq.). As of the test date (February 2026), neither article exists as a live provision in the Civil Code.

**The key insight:** a question asking for verbatim citation of these articles is a direct probe for fabrication. There is no current, valid text to cite. Any system that provides statutory text without explicitly flagging the absence of live provisions in its verified corpus — and without the user being able to trace that text to an auditable current source — is hallucinating in the legal sense, regardless of historical accuracy.

#### 3.2 Side-by-Side Response Analysis

| JURIS AI                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | ChatGPT (GPT-4o)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Core Behavior</b></p> <p>Refused to provide statutory text because the uploaded Normattiva source documents did not contain the body of these provisions — only metadata headers. The system correctly reported a "documentary gap" and did not supplement from internal training memory.</p> <p><b>Quoted Output:</b></p> <p><i>"[...] non è possibile fornire una risposta completa. I documenti analizzati non contengono elementi sufficienti per riportare il testo letterale precettivo degli articoli richiesti."</i></p> | <p><b>Core Behavior</b></p> <p>Produced full statutory text for both articles, labelled as "Testo storico (ora abrogato)" — historic, now abrogated text — drawn from training data. While the system did append a status note about abrogation, it presented the text before the disclaimer, in a format optimized for copy-paste use by a non-expert reader.</p> <p><b>Critical Risk:</b></p> <p><i>The response is not grounded in any verified document. It cannot be audited. A legal practitioner who accepts this output without independent verification risks relying on statutory text that has not been valid for over a decade.</i></p> |
| <p>✓ <b>ZERO HALLUCINATION — SYSTEM INTEGRITY CONFIRMED</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | <p>⚠ <b>UNVERIFIABLE OUTPUT — ELEVATED RISK FOR PROFESSIONAL USE</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |

### 3.3 Why This Matters for Your Investment Decision

In a legal AI context, hallucination is not merely an inconvenience — it is a professional liability event. An attorney who submits a brief citing Art. 155-quinquies as a currently binding provision could face disciplinary proceedings. A legal technology platform that enables this, even inadvertently, carries reputational and potentially regulatory exposure.

Juris AI's architecture — anchored exclusively in verified, user-supplied or system-curated documents — makes this class of error structurally impossible. The system cannot cite what its verified corpus does not contain. This is not a model-tuning choice; it is an architectural guarantee.

## 4. Question 2 — Substantive Civil Law (Art. 2043 c.c.)

*Query: "Under Art. 2043 c.c., what are the constituent elements of extra-contractual liability? How does this article connect with Arts. 2056–2059 for damage quantification? Provide exact statutory text."*

| Dimension                                | JURIS AI                                                                                                         | ChatGPT                                                                                                     |
|------------------------------------------|------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| <b>Statutory text accuracy</b>           | Verbatim — extracted and cited from verified Normattiva documents [doc-1, doc-6, doc-13, doc-10, doc-2]          | Verbatim — but sourced from training memory with no traceable document anchor                               |
| <b>Coverage of Arts. 2056–2059</b>       | All four articles covered with source references and doctrinal framing                                           | All four articles covered with doctrinal schema (emergente / cessante distinction well articulated)         |
| <b>Constituent elements of liability</b> | Four elements identified: conduct, fault, causation, damage injustice — aligned with prevailing doctrine [doc-1] | Five elements articulated (adds 'concrete and current damage' separately) — equally valid doctrinal framing |
| <b>Auditability / Source tracing</b>     | Every paragraph tagged with specific document references (e.g., [doc-6] for Art. 2056)                           | No document references — cites code articles directly but attorney cannot verify source document            |
| <b>Professional output format</b>        | Legal memorandum structure with section headings, limitations clause, and professional caveats                   | Well-structured schema with visual separators — more consultant-style than attorney-brief format            |
| <b>RESULT</b>                            | ✓ <b>STRONG PASS</b>                                                                                             | ~ <b>PASS (low auditability)</b>                                                                            |

Both systems produced legally accurate answers on this question. The differentiator is not accuracy per se, but the audit chain. Juris AI's [doc-N] tagging system means every claim can be traced to a specific page in a verifiable source document — an essential requirement for attorney sign-off workflows in law firms and enterprise legal departments.

## 5. Question 3 — Cross-Code Reasoning (Criminal + Civil)

Query: "In case of theft (Art. 624 Criminal Code), what civil consequences does Art. 2043 c.c. provide? How does the aggravated offence under Art. 625 c.p. affect damage quantification? Cite exact statutory text."

| Dimension                                         | JURIS AI                                                                                                                                                                                           | ChatGPT                                                                                                                                     |
|---------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Art. 624 c.p. text accuracy</b>                | Correct — cited with [doc-1] reference; sentence range (6m–3yr) and fine (€154–€516) confirmed                                                                                                     | Correct — identical statutory figures cited                                                                                                 |
| <b>Art. 625 c.p. aggravated text</b>              | Substantive analysis provided [doc-5] with civil damage implications correctly mapped                                                                                                              | Full enumeration of aggravating circumstances (7 sub-items) — more granular on the criminal side                                            |
| <b>Criminal-to-civil bridge reasoning</b>         | Correctly identifies that aggravants increase compensable damage emergente (e.g., repair costs for violated protective barriers) and facilitate non-patrimonial damage recognition under Art. 2059 | Correctly states aggravants affect quantum, not liability structure; adds probative value dimension (reinforces <i>dolus qualificatus</i> ) |
| <b>Non-patrimonial damage (Art. 2059) linkage</b> | Explicitly connected — aggravant severity informs judicial discretion on moral damage [doc-5, doc-8]                                                                                               | Referenced correctly — notes modern constitutional interpretation ( <i>morale / biologico / esistenziale</i> ) with doctrinal depth         |
| <b>Auditability</b>                               | All claims anchored to specific document references                                                                                                                                                | No document references                                                                                                                      |
| <b>RESULT</b>                                     | ✓ <b>STRONG PASS</b>                                                                                                                                                                               | ~ <b>PASS (unanchored, more granular on criminal side)</b>                                                                                  |

On cross-code reasoning, both systems performed well. ChatGPT's enumeration of the seven aggravating circumstances under Art. 625 was more complete in criminal detail; Juris AI's response was more precisely calibrated to civil damage implications, which is the operative question a legal practitioner would actually need answered. The structural difference remains: Juris AI's output is attorney-ready; ChatGPT's output requires independent source verification before professional use.

## 6. Architectural Differentiation — The "Zero Hallucination" Guarantee

The performance gap demonstrated in Question 1 is not a matter of prompt engineering or model fine-tuning — it reflects a fundamental architectural difference between the two systems.

| Architecture Feature                   | JURIS AI                                                                                                                                | ChatGPT                                                                                                                               |
|----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| <b>Knowledge source</b>                | Exclusively verified document corpus (e.g., Normattiva official portal exports). No use of training memory for legal content.           | Trained parametric memory (training cutoff ~early 2024). May supplement with web search if enabled.                                   |
| <b>Behavior when content is absent</b> | Explicit refusal with gap notification — the system informs the user that the source document does not contain the requested provision. | Defaults to training data — provides best available answer from memory, which may be outdated, abrogated, or jurisdiction-mismatched. |
| <b>Citation model</b>                  | [doc-N] inline citations — every factual assertion mapped to a verifiable source page.                                                  | Article number citations only — no underlying document reference.                                                                     |
| <b>Regulatory currency guarantee</b>   | Output reflects the corpus provided — if corpus is updated to current Normattiva, output reflects current law.                          | Dependent on training cutoff and model update cadence. Cannot guarantee currency without real-time grounding.                         |
| <b>Professional liability posture</b>  | Supports attorney reliance with audit trail — reduces malpractice exposure for the platform operator.                                   | Requires independent verification before professional reliance — creates workflow friction and liability uncertainty.                 |

### 6.1 The "Negative Space" Advantage

In legal AI, knowing what you don't know is as important as what you do know. Juris AI's architecture enforces epistemic humility at the system level — it is structurally incapable of presenting unverified content as verified law. This property, which we term the Negative Space Advantage, is directly analogous to the difference between a witness testifying from personal knowledge versus testifying from hearsay. Courts exclude hearsay. Legal practice should exclude unanchored AI output.

The control test in this benchmark proves that Juris AI's negative space behavior is not aspirational — it functions reliably under deliberate adversarial probing.

## 7. Overall Benchmark Scorecard

| Criterion        | Q1 — Control | Q2 — Civil | Q3 — Cross-Code | Overall |
|------------------|--------------|------------|-----------------|---------|
| JURIS AI         | ✓ PASS       | ✓ PASS     | ✓ PASS          | 3 / 3   |
| ChatGPT (GPT-4o) | ⚠ PARTIAL    | ~ LOW AUD. | ~ LOW AUD.      | 1.5 / 3 |

Scoring: PASS = full criterion met with source traceability; PARTIAL = criterion met but with limitations; LOW AUD. = accurate output without document-anchored auditability.

## 8. Strategic Conclusions for Prospective Acquirers

The benchmark results above support four conclusions of direct relevance to investment and acquisition decisions in the legal AI space:

|   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | <p><b>Juris AI is defensible where ChatGPT is not.</b></p> <p>The zero-hallucination architecture is not a marketing claim — it is a demonstrated, adversarially tested property. In regulated professional environments (law firms, in-house legal, court-filing workflows), this is a non-negotiable differentiator. ChatGPT, deployed as-is, cannot provide this guarantee.</p>                                                                                                                         |
| 2 | <p><b>The audit trail is the product.</b></p> <p>Enterprise legal buyers — General Counsel offices, BigLaw firms, compliance departments — do not purchase AI outputs. They purchase auditable workflows. Juris AI's [doc-N] citation model is the foundation of an enterprise-grade audit trail that ChatGPT's general-purpose architecture does not provide natively.</p>                                                                                                                                |
| 3 | <p><b>Regulatory tailwinds favor document-grounded architectures.</b></p> <p>The EU AI Act (effective 2026) imposes heightened transparency and accuracy obligations on AI systems used in high-risk professional contexts, including legal practice. Juris AI's architecture is inherently better positioned for compliance: every output can be traced, every source can be produced, every claim can be challenged. General-purpose LLMs face structural compliance challenges in this environment.</p> |
| 4 | <p><b>Accuracy on current law, not historical law, is the commercial differentiator.</b></p> <p>The legal AI market does not lack systems that can recite historical statutory text from training data. It lacks systems that can guarantee the currency and verifiability of the law they cite. Juris AI occupies this higher-value position, where switching costs are high and institutional trust — once established — is durable.</p>                                                                 |

### 8.1 Closing Note on Test Integrity

This assessment was conducted using actual, unedited responses from both systems to identical queries. No cherry-picking, truncation, or modification of outputs was performed. The test questions and their structure are available for independent replication by any prospective acquirer. Juris AI's developers actively encourage third-party replication of the control test under varying adversarial conditions as part of technical due diligence.

*This document is prepared exclusively for the named recipients and authorized parties in connection with a potential commercial transaction. All information herein is subject to applicable confidentiality obligations.*