

The State of Agent Verification 2026

Five requirements for verifiable agent action,
assessed for legal practice.



In brief

As lawyers seek to use agents for more and more tasks, they need to be able to prove their own rules were applied to control an agent before an action executed. This report defines five requirements for verifiable agent action, maps the legal duties that already impose them, and grades current controls against the standard.

THE SHIFT

Agents now act. The constraint moved from capability to trust.

03

Adoption is near-universal, and fabricated legal-citation cases grew nearly 20x in fourteen months.

THE LEGAL DUTY

The duty to control agents predates AI.

04

Half a century of law already requires controlled, provable automated action.

THE DEFINITION

Verifiable agent action has five requirements.

06

Pre-action, deterministic, independently verifiable, portable, and workable without exposing the rules or the data.

THE DISTINCTION

Logs assert. Receipts prove. The difference is evidence an outside party can verify.

07

A log is the system's account of itself. Checking a receipt requires no trust in the operator.

THE ASSESSMENT

Every approach is strong somewhere, but each misses something.

08

No approach family meets all five requirements except enforcement with proof.

THE WORKED EXAMPLE

Meeting all five is possible today.

09

Plain-English rules, a solver verdict before the action runs, and a receipt anyone can verify in under a second.

WHAT HAPPENS NEXT

Three moves for the market.

10

Buyers put the five questions in the RFP. Builders publish receipts, not accuracy claims. Standards bodies get a starting point for a neutral definition.

Agents now act. The constraint moved from capability to trust.

The question in front of lawyers is narrow and hard: how much work can a lawyer responsibly hand to software that acts on its own.

The market has answered what the software can do. Agents at plaintiff firms open insurance claims and sign retainers on intake calls. AI-native firms draft and serve discovery responses. In the UK, Garfield Law, the first SRA-authorized firm delivering legal services through AI, issues letters before action and files claims.

Supervision did not keep pace. ABA Formal Opinion 512 and Model Rule 5.3 keep the duty on the lawyer, not the software. A lawyer running agents at scale is making a promise about supervision they usually cannot document, and the sanctions record shows the promise is often broken. Most of the sanctioned firms had AI policies in place when it happened but they lacked the ability to enforce them.

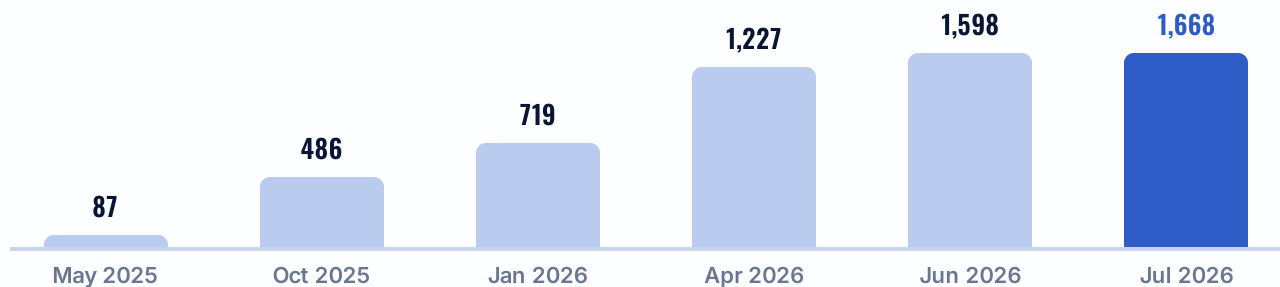
79%

of lawyers now use AI in their practice.

Source: ABA 2026 TechReport.

EXHIBIT 1

Fabricated legal-citation cases grew nearly 20x in fourteen months.



Note: Cumulative court decisions worldwide in which a court found, or clearly implied, reliance on AI-hallucinated material. Counts are snapshots on the dates shown; the database adds cases near-daily, so any figure is a floor.

Source: Damien Charlotin, AI Hallucination Cases database, counts as reported May 2025 to July 2026; ICME Labs analysis.

The failures trace to citation checks that never ran, not to bad checking tools.

That distinction is the subject of this report. What has not improved is the ability to guarantee, and then to prove, that an organization's own rules were applied to a specific agent action **before it executed**. That missing ability is the verifiability gap, and it now sets the ceiling on what lawyers allow agents to do.

The duty to control agents predates AI.

Each of the authorities below imposes some combination of the same three obligations on automated action: control it before it executes, prove the control operated, and let someone outside the operator confirm it. Agent verification is a fifty-year-old duty applied to a new actor.

EXHIBIT 2

Half a century of law already requires controlled, provable automated action.

	AUTHORITY	CONTROL BEFORE THE ACTION	OPERATOR PROVES IT OPERATED	OUTSIDE PARTY CAN VERIFY
1977	IEEPA / OFAC sanctions Screen and block prohibited transactions before they settle; liability attaches regardless of intent	●	○	●
1989	UCC Article 4A Commercially reasonable security procedures applied to payment orders before acceptance	●	●	●
1997	21 CFR Part 11 Electronic records must carry secure, computer-generated audit trails that detect alteration	○	●	●
2010	SEC Rule 15c3-5 Automated pre-order risk controls on market access, certified annually by the CEO	●	●	●
2017	NYDFS Part 504 Transaction filtering and watchlist programs, certified to the regulator by the board or a senior officer	●	●	●
2017	FRE 902(14) Electronic records verified by hash value are self-authenticating in evidence		●	●
2018	GDPR Article 5(2) The controller is responsible for compliance and must be able to demonstrate it	○	●	●
2024	ABA Op. 512 · Rules 5.3, 1.15 Lawyers must supervise nonlawyer assistance, including AI tools, and safeguard client funds	●	○	○
2026	EU AI Act Article 12 High-risk AI systems must automatically record events across their lifetime	○	●	●

● Explicit requirement ○ Implied or partial

Note: Characterizations are summaries of each authority's control and evidence obligations, not legal advice; scope and applicability vary by actor and activity.

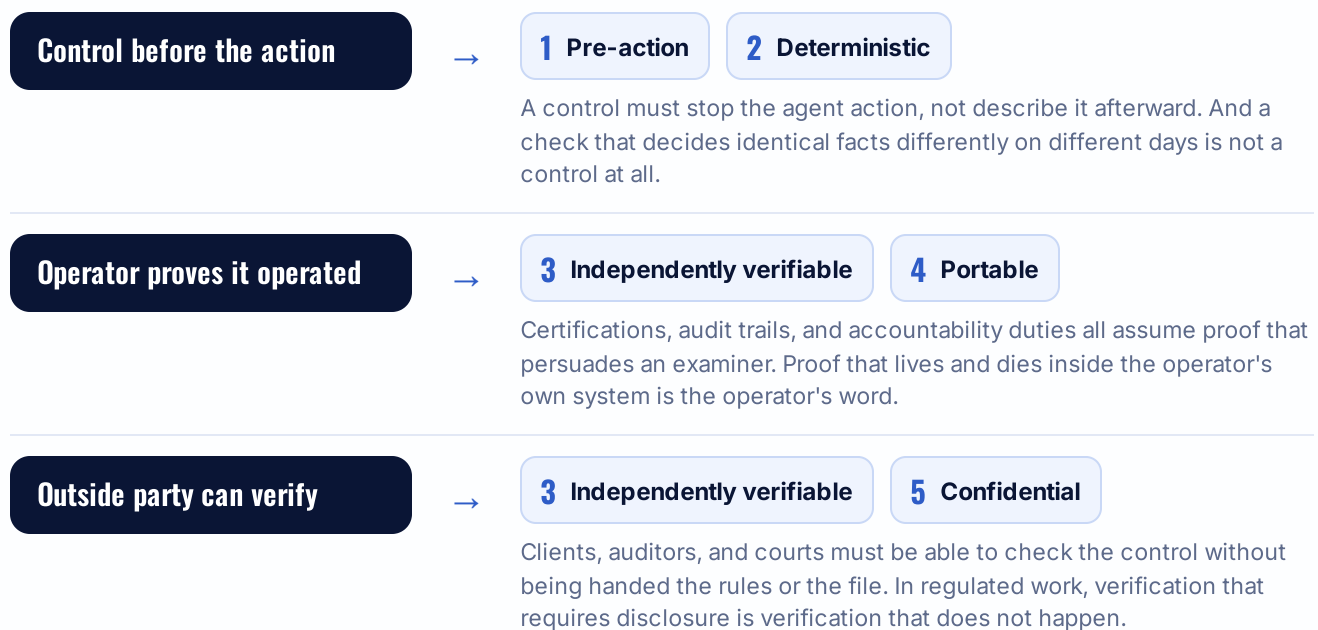
Source: Cited statutes, regulations, rules, and opinions; ICME Labs analysis.

Five requirements for agents meet what the law already asks.

The core obligations from Exhibit 2 have applied to automated action for fifty years. The next section restates them as five requirements for agents.

EXHIBIT 3

Three legal obligations expand into five requirements for agents.



Source: Authorities in Exhibit 2; ICME Labs analysis.

The evidence rules already anticipate the receipt. Federal Rule of Evidence 902(14) treats electronic records verified by hash value as self-authenticating: on a written certification, the record comes into evidence without a live witness to vouch for it. A cryptographic receipt is engineered to be that kind of record.

NIST AI RMF

The govern and measure functions call for documented, demonstrable risk controls, not asserted ones.

ISO/IEC 42001

AI management systems certify on auditable evidence that controls operate as described.

OWASP Agentic Security

The 2026 threat framework for agentic systems; its mitigations concentrate before the action runs.

The risk-management frameworks converge from the other direction: a control you cannot evidence is treated as a control you do not have. The next section states the five requirements as a single test.

Verifiable agent action has five requirements.

Each requirement names a property a control must have in practice.

- 1 The check runs before the agent action.**
A control that flags a bad wire transfer after the money moved is a log, not a safeguard. Verification has to happen while the action can still be stopped.
- 2 The verdict is deterministic.**
The same agent action, checked against the same rule, must return the same answer every time. A check that can clear an action on Monday and block the identical action on Tuesday proves nothing.
- 3 Someone outside the system can verify it.**
If only the vendor or the operator can confirm the check happened, the control is self-attested. A client, an auditor, or a regulator must be able to confirm it independently.
- 4 The proof travels.**
Verification locked inside one platform dies there. The evidence must be checkable outside the system that produced it, by someone who has no account on it.
- 5 It works without exposing the rules or the data.**
In regulated work the rule itself and the client file are often confidential. Verification that requires disclosing them is unusable where it matters most. This requirement binds when the work is sensitive, which in legal practice is the default.

KEY TAKEAWAY

These five requirements are demanding. A control meeting only two or three of them can be useful, but a control meeting all five is the evidence the law already asks for.

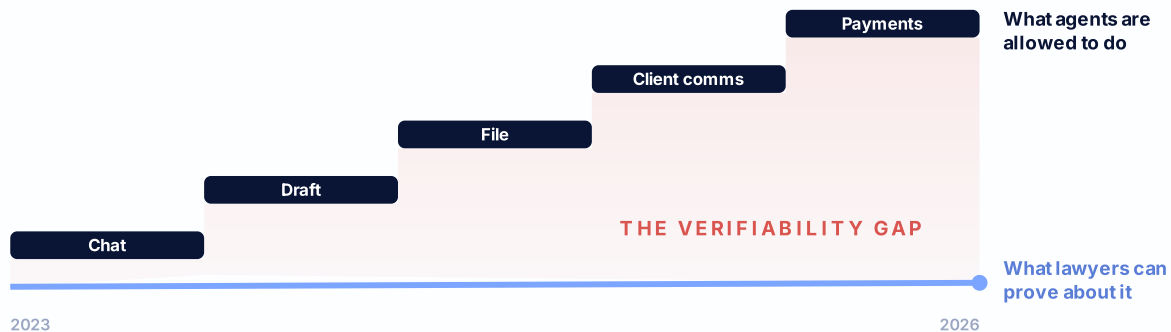
THE DISTINCTION

Logs assert. Receipts prove. The difference is evidence an outside party can verify.

A receipt provides much stronger evidence for agent compliance than a mere log.

EXHIBIT 4

The gap in verifiability is increasing as agents take on more capabilities.



Note: Conceptual illustration. Milestones represent the typical scope expansion of deployed legal agents, 2023 to 2026; vertical position is ordinal, not measured.

Source: ICME Labs analysis.

EXHIBIT 5

Four properties separate a log from a receipt.

System log

```
14:52:07 filing_check status=passed  
user=agent_7 doc=brief_218.pdf
```

- **Self-authored.** Written by the system about its own behavior.
- **After the fact.** Records the action once it has already run.
- **Editable.** An administrator can change it, and the audit trail of the change lives in the same system.
- **Trust required.** Persuasive only to someone who already trusts the operator.

Verification receipt

```
UNSAT · rule v3.2 · filing blocked ·  
receipt #RCPT-7F3A · verify at any browser
```

- **Pre-execution.** The verdict exists before the agent action runs, and a failed check stops it.
- **Rule-bound.** Names the exact rule and version it was decided against.
- **Tamper-evident.** Any alteration is detectable by anyone who checks.
- **No account needed.** A client, carrier, or auditor verifies it without trusting, or even contacting, the operator.

Note: Property comparison of a self-attested system log and a cryptographic verification receipt; example entries are illustrative.

Source: ICME Labs analysis.

Every approach is strong somewhere, but each misses something.

Four families of controls dominate the market today. Exhibit 6 grades each against the five requirements.

EXHIBIT 6

No approach family meets all five requirements except enforcement with proof.

Approach	1 Pre-action	2 Deterministic	3 Independent	4 Portable	5 Confidential
Logging & monitoring	 After the fact	 Records stable	 Self-attested	 Exportable	 Varies
LLM-judge review	 Can gate actions	 Probabilistic	 Not reproducible	 Host-dependent	 Judge sees data
Formal checks, post-inference e.g. AWS Automated Reasoning	 Post-inference	 Formally verified	 Within one cloud	 Platform-bound	 Operator-visible
Signed attestation pre-action authorization, signatures	 Gates the action	 Rule-based	 Signature verifies	 Travels	 May expose rules or data
Enforcement with proof solver verdict + cryptographic receipt (ICME Preflight)	 Blocks before execution	 Solver, not a model	 Anyone can verify	 Receipt travels	 No rule or file exposure

 Meets the requirement
  Partially or conditionally
  Does not meet it

Note: Assessments reflect publicly documented behavior of each approach family; vendor implementations vary. Grade any specific product, including ours, against the five requirements directly.

Source: Public vendor documentation as of July 2026; ICME Labs analysis.

The table shows a young field with real strengths distributed across it. Formal post-inference checks solved determinism. Signed attestation solved portability. What no approach solved together, until the last row, is all five at once, and the fifth column is where the difference concentrates: **proving a check ran without exposing what was checked.**

Meeting all five is possible today.

ICME Labs built Preflight to meet this standard, and the mechanics are simple to state.

1. The lawyer writes the rules in plain English.
2. The rules compile once into formal logic.
3. Before an agent acts, a mathematical solver checks the proposed agent action against the rules.
4. The solver returns a verdict: SAT, the action complies and may run, or UNSAT, it does not and is blocked.

Every decision produces a tamper-evident cryptographic receipt. A client, a malpractice carrier, or a bar auditor can verify it in under a second, without access to the rules or the client's file. The receipt renders as a plain-language report a lawyer reads in ten seconds, backed by the cryptographic proof a machine can check underneath.

The report is what you read. The receipt is what you cannot fake.

The receipt proves the rule fired against the facts the agent asserted. It does not prove those facts are true, and it does not make a filing correct. What it removes is a previously unanswerable question: whether the check ran at all.

Lawyers who can prove their rules held on a filing, a disbursement, or a client communication can put agents on filings, disbursements, and client communications. Enforcement first, then proof, then scope.

The report below is a simulation of the check this section describes. An agent attempts a court filing without the required citation check. The rule fires, Preflight blocks the filing, and the verification report records the block.

EXHIBIT 7

The report is readable in ten seconds. The proof underneath is machine-checkable.



Note: Illustrative simulation. An agent attempts a court filing without the firm's required citation check; the rule fires and the filing is blocked.

Source: ICME Labs Preflight verification report rendering.

Three moves for the market.

For buyers. Put the five requirements in the RFP. Ask any vendor of agent controls, including ICME, where the check runs, whether the verdict is repeatable, who outside the system can confirm it, whether the proof travels, and whether verifying requires disclosure. The questions take five minutes and they separate evidence from assertion faster than any demo.

For builders. Publish receipts, not accuracy claims. The field's habit of self-graded benchmarks is the same structural problem as self-attested logs. An artifact a skeptic can verify beats a percentage a vendor reports.

For standards bodies. The five requirements are offered here as a starting point for a vendor-neutral definition of verifiable agent action. A standard maintained by a neutral body, against which any implementation can be graded, would serve buyers better than any one company's framing, including ours.

Agents will take on more of the work of law. Whether they act inside lawyers' own rules or around them turns on evidence, not on models. Evidence has a definition.

About the author. Houman Shadab is co-founder of ICME Labs and a Fellow at Stanford CodeX. He is a widely influential and award-winning legal scholar, three-time Congressional witness, and spoke by invitation at the Federal Reserve Bank of New York's 2026 Innovation Conference on turning law into verified code. Contact: houman@icme.io.

About ICME Labs. ICME builds Preflight, pre-action verification for AI agents: plain-English rules, deterministic enforcement, and cryptographic receipts anyone can verify. icme.io

