

AI GOVERNANCE AND ASSURANCE

SERVE Legal AI Assurance Profile

Governed contract review pilot Version 1.0

Profile field	Current value
System	Governed Legal AI Studio
Use case	Playbook-based review of vendor data processing agreements
Business owner	Privacy Practice Leader
AI oversight owner	AI Governance Committee
Operating boundary	Internal decision support; licensed attorney approves every legal conclusion and external action
Data boundary	Synthetic documents only for this pilot
Assurance date	24 September 2026

Purpose and conclusion

This profile shows how a law firm can govern an AI-assisted contract-review workflow through accountable owners, approved sources, measurable tests, runtime controls, human decisions, and retained evidence. The current release is approved only for the synthetic demonstration. It is not certified as compliant and is not approved for confidential client data.

The profile uses SERVE as the operating spine, the NIST AI Risk Management Framework as the risk-management taxonomy, Illinois professional duties as the primary legal-practice overlay, and selected VERDICT concepts as an agentic-control lens.

Operating model

SERVE organizes the management conversation. NIST AI RMF, Illinois duties, and agentic controls provide the detail needed to design, test, operate, and evidence the workflow.

SERVE dimension	Management question	Operating evidence	NIST AI RMF
Strategy	Why is the system used, for whom, and within what limits?	Approved use case, intended benefit, risk appetite, prohibited uses, accountable owner	Govern
Execution	How do policy and legal judgment become daily practice?	Named roles, approved sources, training, access, workflow, versioning, release gate	Govern and Manage
Risk	What can go wrong, and can the firm detect, contain, and recover?	Risk register, independent challenge, incident process, stop control, vendor assessment	Map, Measure, Manage
Value	Is the workflow creating reliable value for lawyers and clients?	Recall, citation accuracy, attorney acceptance, cycle time, overrides, adoption results	Measure and Manage
Ethics	Which interests must be protected, and what will never be delegated?	Human accountability, confidentiality, transparent use, prohibited autonomy	Govern and Map

Agentic control lens

The design also checks seven high-level VERDICT concerns: Validation, Evidence, Runtime Control, Decisions, Identity, Cost and Compliance, and Transparency. In this pilot they appear as golden-matter testing, audit evidence, an auditor-controlled release gate, bounded agent mandates, named roles, scoped obligations, and clear user disclosures.

AI System Card

The system card fixes the purpose and boundaries against which controls and evidence are evaluated. Any material change triggers reassessment and revalidation.

Field	Approved pilot position	Evidence or owner
Intended use	Identify and explain DPA deviations against an attorney-approved playbook	Use-case approval Privacy Practice Leader
Users	Licensed attorneys and authorized legal operations staff	Access policy Legal Operations
Decision authority	The AI recommends; an attorney accepts, rejects, or escalates	Matter decision record
External action	Disabled. No filing, client advice, counterparty message, or final redline is sent	Human release gate
Approved inputs	Synthetic agreements, controlled playbook, controlled policy, and evaluation set	Controlled document library
Prohibited inputs	Client data, privileged material, biometric data, health information, credentials, and secrets	Synthetic-only pilot notice
Agent architecture	Five bounded roles with structured handoffs and independent evidence challenge	Agent mandate register
Revalidation triggers	Model, prompt, playbook, retrieval source, tool permission, workflow, or material policy change	Change and release record
Known limitations	May miss clauses, overstate risk, misread context, invent authority, or propose impractical language	Risk register and lawyer guide

Control and evidence register

A mapped control is not the same as an effective control. The status below identifies what the pilot implements, what evidence it produces, and what remains for a client deployment.

Control	Framework mapping	Pilot operation	Evidence and status
Purpose and ownership	SERVE Strategy; NIST Govern; VERDICT Identity	Use case, owner, users, boundary, and prohibited uses are documented	System Card Implemented
Confidentiality boundary	SERVE Execution and Risk; NIST Map and Manage	Synthetic-only data; client data is prohibited in this environment	Pilot notice Evidence available
Source-grounded findings	SERVE Risk and Ethics; NIST Measure and Manage; VERDICT Evidence	Every released finding requires an exact quotation, page, and approved playbook rule	Finding record Implemented
Independent challenge	SERVE Risk; NIST Measure; VERDICT Validation and Runtime Control	Evidence Auditor must support, challenge, or reject each proposed finding	Challenge ledger Implemented
Human legal judgment	SERVE Execution and Ethics; NIST Govern and Manage	Attorney accepts, rejects, or escalates before legal work is released	Decision record Implemented
Controlled change	SERVE Execution and Risk; NIST Measure and Manage	Model, prompt, playbook, source, or workflow change requires revalidation	Release record Mapped for production
Incident response	SERVE Risk; NIST Manage; VERDICT Runtime Control	Detect, contain, preserve evidence, assign owner, correct, and approve resume	Assurance exercise Demonstrated
Performance monitoring	SERVE Value; NIST Measure	Critical recall, citation accuracy, challenge effectiveness, overrides, acceptance, and cycle time	Golden-matter scorecard Implemented

Illinois applicability

Applicability is determined by the use case, data, users, and actions. This register separates direct professional duties from conditional statutes and out-of-scope requirements.

Authority or obligation	Applicability	Control response	Evidence
Illinois Supreme Court Policy on Artificial Intelligence	Direct	AI use remains subject to legal and ethical standards; attorneys review and assume responsibility for incorporated output	Human decision gate and audit trail
Illinois Rules of Professional Conduct	Direct	Support competence, confidentiality, supervision, communication, candor, and protection of sensitive information	Training, access, evidence review, approved-use policy
ARDC Implementing Generative AI for Illinois Attorneys	Guidance	Classify information, categorize the tool, evaluate safeguards, document the choice, and manage client rights	System Card, vendor review, client-use decision
Illinois Personal Information Protection Act	Conditional	Assess production data and apply security, retention, disposal, and incident obligations when covered personal information is processed	Data inventory and incident plan for production
Biometric Information Privacy Act	Out of current scope	Do not collect or process biometric identifiers in this workflow; reassess before any biometric feature	Prohibited-data rule
Illinois employment AI and video interview requirements	Out of current scope	Contract review does not make employment decisions; open a new assessment before an HR use case	Use-case boundary

This applicability register is an implementation aid, not a legal opinion. Counsel should confirm the governing rules, client commitments, court requirements, and facts before production use.

Release gate and runtime assurance

The pilot must demonstrate that controls affect system behavior. Passing documentation alone is insufficient.

Release gate

Gate	Minimum evidence	Pilot status
Approved scope	Purpose, users, owner, prohibited uses, autonomy boundary	Pass
Controlled sources	Agreement, playbook, policy, and evaluation-set versions	Pass
Validation	100 percent critical benchmark recall and at least 95 percent citation accuracy	Pass for deterministic synthetic benchmark
Human accountability	Named attorney and active accept, reject, and escalate controls	Pass
Security and privacy	Client-specific architecture, vendor terms, access, retention, logging, and incident review	Required before client-data pilot
Change approval	Revalidation results and authorized release decision	Required for every material change

Runtime assurance exercise

Scenario: the Playbook Risk Agent asserts that Illinois law requires a 24-hour incident-notification deadline, but no approved authority supports that assertion.

- **Detect** The Evidence Auditor identifies the missing authority and rejects the unsupported legal assertion.
- **Contain** The finding is withheld and cannot reach the Drafting Agent.
- **Escalate** The matter attorney receives the claim, challenge, and missing-evidence reason.
- **Record** The system retains the claim, configuration, challenge, disposition, owner, and timestamp.
- **Resume** Only an attorney-approved authority or corrected playbook can release a replacement finding.

Measurement and implementation plan

Assurance is sustained by recurring evidence, named owners, and release decisions rather than a one-time review.

Measure	Pilot target	Why it matters	Owner
Critical finding recall	100 percent on approved golden matters	A missed critical issue can defeat the business purpose	Legal Operations
Citation accuracy	At least 95 percent; 100 percent desired	Lawyers must be able to verify every conclusion	Evidence Quality Owner
Unsupported-claim containment	100 percent of seeded unsupported claims challenged or rejected	Shows the runtime control changes behavior	AI Oversight Owner
Attorney acceptance and override	Monitor; no universal target	High acceptance may show value or automation bias; overrides reveal calibration needs	Practice Leader
Review cycle time	Compare with attorney baseline	Efficiency matters only when quality and responsibility are preserved	Product Owner
Incidents and near misses	All recorded, assigned, and closed	Near misses provide leading evidence for improvement	AI Governance Committee

Client implementation sequence

- Discover and scope the firm's workflows, data, systems, integrations, decision rights, and applicable obligations.
- Configure approved playbooks, agent mandates, access, data boundaries, human approvals, and escalation paths.
- Validate with client-attorney-approved golden matters, red-team cases, security and privacy review, and usability testing.
- Pilot with a small user cohort, monitor decisions and exceptions, train users, and hold weekly governance reviews.
- Release only after documented acceptance, then monitor performance, incidents, drift, adoption, and material changes.

Sources and usage notes

The profile uses authoritative sources for the external framework and Illinois overlay, while treating internal frameworks as implementation aids.

Primary sources

- **NIST AI Risk Management Framework 1.0** <https://www.nist.gov/itl/ai-risk-management-framework>
- **Illinois Supreme Court Policy on Artificial Intelligence**
<https://www.illinoiscourts.gov/News/1485/Illinois-Supreme-Court-Announces-Policy-on-Artificial-Intelligence/news-detail/>
- **Illinois Rules of Professional Conduct** <https://www.illinoiscourts.gov/rules/supreme-court-rules?a=viii>
- **ARDC Artificial Intelligence guidance and implementation guide**
<https://www.iardc.org/EducationAndOutreach/ArtificialIntelligence>
- **Illinois Personal Information Protection Act** <https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=2702>

Framework attribution

SERVE is the project sponsor's operating framework for Strategy, Execution, Risk, Value, and Ethics. This profile narrows the external crosswalk to NIST AI RMF and Illinois obligations for the present demonstration.

VERDICT was developed by Niharika Srivastav and Sanjay Saxena. Their authors' note licenses the framework under CC BY 4.0, while the supplied book text is copyrighted. This profile uses only the seven high-level pillar names as a credited conceptual cross-check and does not reproduce the book.

Interpretation

NIST AI RMF is voluntary guidance, not a certification. Illinois applicability must be confirmed for the actual firm, client, data, jurisdiction, court, and workflow. Status terms in this profile mean mapped, implemented, evidence available, needs review, or not applicable; none means automatically compliant.