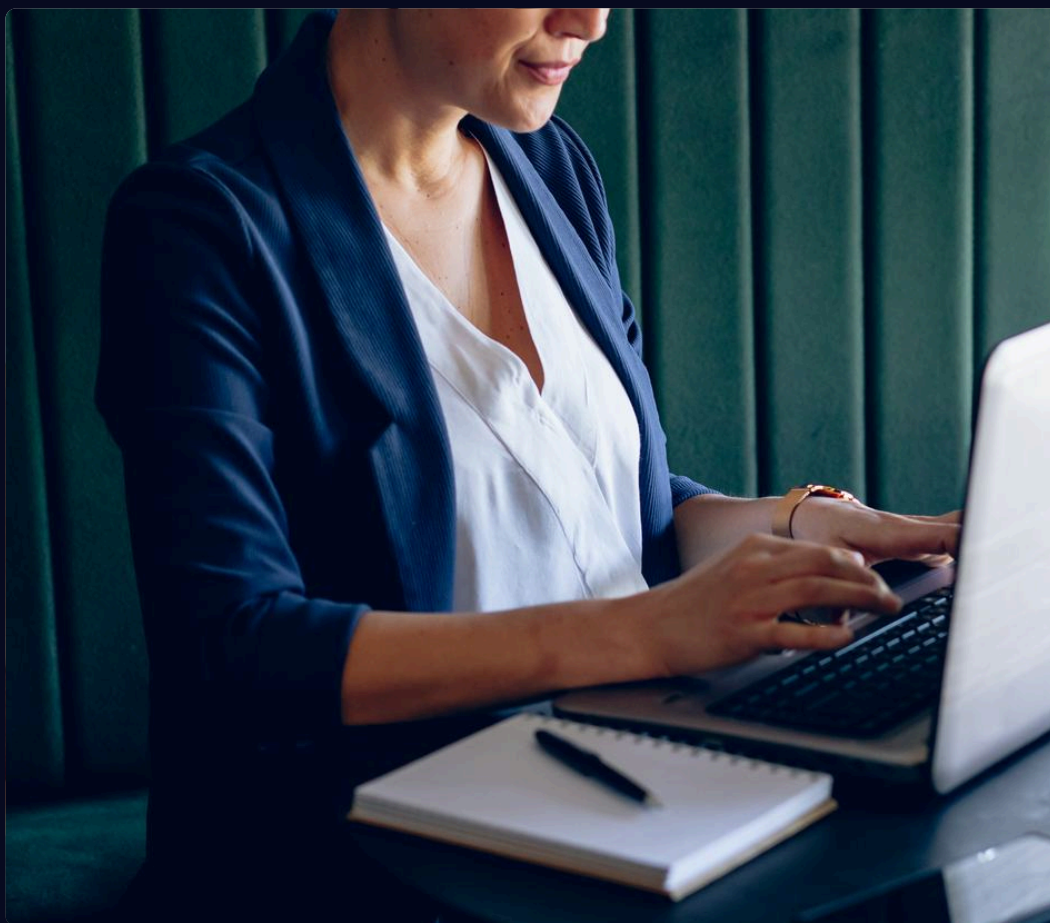# alexi

Evaluating Legal AI Responsibly

A Framework for Law Firm Decision-Makers

A guide to help firms assess AI tools for quality, confidentiality, and alignment with professional standards.

Purpose of This Framework

The adoption of legal AI carries professional obligations as well as operational promise. Evaluating tools solely on speed or novelty risks overlooking what truly determines success in practice: accuracy, confidentiality, consistency, and control.

This framework is designed to help firm leadership, IT, and innovation teams assess prospective solutions in a structured, balanced way, focusing on what matters most to legal professionals and their clients.
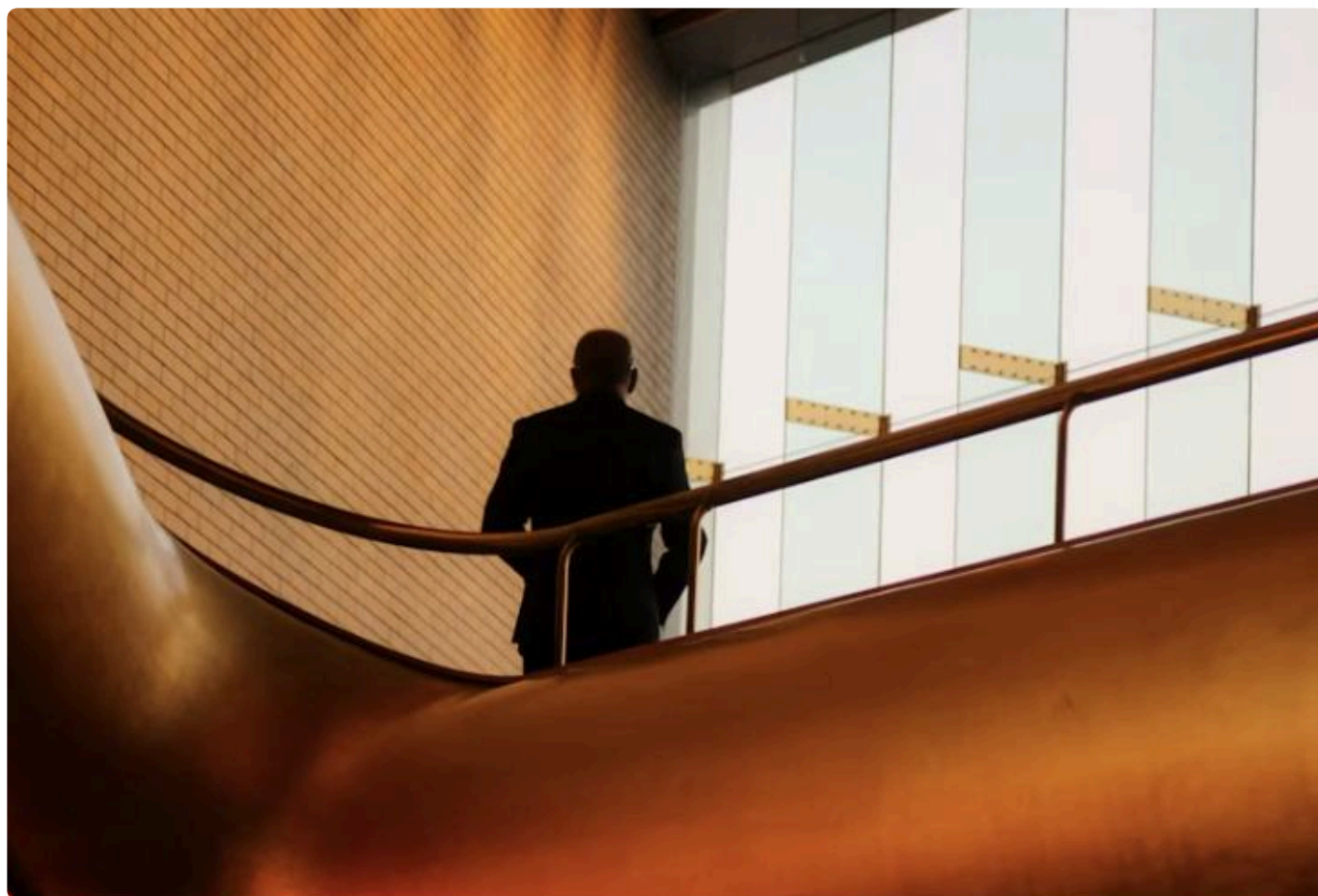
01 Legal Accuracy and Reasoning

What to Examine:
- Does the system reason in a way lawyers can follow and verify?
- Are its analyses grounded in cited sources: statutes, regulations, or precedent?
- Can lawyers review, challenge, and refine its output?

Why It Matters:
Precision and transparency are the foundation of credible legal work. Tools that cannot explain or evidence their reasoning introduce professional risk.

Indicators of Maturity:
Produces cited, verifiable output consistent with firm drafting standards and preserves a clear review trail.

02 Confidentiality and Data Control

What to Examine:
- How and where is client and firm data processed and stored?
- Can the firm control data retention, usage, and access?
- Is the environment private to the firm, or shared with others?

Why It Matters:
Client confidentiality is non-negotiable. Any system that transmits or learns from client data outside firm control may breach professional or client obligations.

Indicators of Maturity:
For the highest level of security, use a privately deployed instance with full encryption, access controls, and firm-managed data policies.

03 Alignment with Firm Standards

What to Examine:
- Does it reflect your firm's templates, tone, and conventions?
- Can it replicate your internal approval and review process?
- Are outputs consistent across lawyers and offices?

Why It Matters:
AI should reinforce the firm's collective standard, not individual variation. Consistency safeguards reputation and quality control.

Indicators of Maturity:
Adapts to firm templates and style, produces review-ready work requiring minimal reformatting.

04 Breadth of Legal Application

What to Examine:
- Does it support both litigation and transactional work?
- Can it assist with drafting, analysis, and review across matter types?
- Is it capable of evolving with new workflows?

Why It Matters:
The right solution supports the full practice, from motions and pleadings to contracts and diligence, ensuring consistent quality across disciplines.

Indicators of Maturity:
Demonstrated effectiveness across multiple practice areas and document types.

05 Ease of Use and Lawyer Adoption

What to Examine:

- Can a lawyer start using it without training?
- Is the interface intuitive and conversational?
- Does it fit naturally into existing document and matter workflows?

Why It Matters:

Adoption depends on familiarity and trust. Tools that feel foreign or require technical skills rarely see meaningful usage.

Indicators of Maturity:

Plain-language interface, minimal onboarding, integrates with DMS and matter systems.

06 Knowledge Retention and Institutional Learning

What to Examine:

- Can internal knowledge and precedents be reused or surfaced?
- Does the system improve through firm use without exposing data externally?
- How easily can knowledge be captured and shared?

Why It Matters:

Firms should retain the value of their own work product. AI should strengthen institutional memory, not disperse it.

Indicators of Maturity:

Secure reuse of internal knowledge, enabling the firm's collective intelligence to grow over time.

07 Ethical and Regulatory Compliance

What to Examine:
- Does the tool comply with emerging AI regulations (e.g., EU AI Act, U.S. AI Bill of Rights, ABA guidance, Canada's proposed Artificial Intelligence and Data Act [AIDA])?
- Are bias, fairness, and explainability assessments performed regularly?
- Does the system allow for human oversight and professional judgment to guide or override automated outputs?

Why It Matters:
Legal AI must not only comply with data and professional conduct standards, but also preserve human oversight, fairness, and accountability. Firms are increasingly expected to demonstrate diligence in how they govern AI use.

Indicators of Maturity:
Regular third-party audits, bias and explainability testing, and formal governance protocols ensuring human oversight remains integral.

08 Partnership, Training and Support

What to Examine:
- Does the provider offer governance support, training, and change management?
- How transparent are they about model behavior and updates?
- Is there a defined process for oversight and escalation?

Why It Matters:
Reliability comes from partnership, not software alone. Continuous collaboration ensures the system remains aligned with firm needs and professional standards.

Indicators of Maturity:
Structured support program, transparent development practices, and clear accountability model.

Evaluation Matrix

Category	Weight	Vendor A	Vendor B	Vendor C
Legal Accuracy and Reasoning	25%	☐	☐	☐
Confidentiality and Data Control	20%	☐	☐	☐
Alignment with Firm Standards	15%	☐	☐	☐
Breadth of Legal Application	15%	☐	☐	☐
Ease of Use and Lawyer Adoption	10%	☐	☐	☐
Knowledge Retention and Institutional Learning	10%	☐	☐	☐
Partnership, Training and Support	5%	☐	☐	☐

Note: These categories reflect the priorities common to most firms, balancing legal quality, data integrity, and long-term value. Weightings can be adjusted to reflect your firm's focus areas.

How to Use This Framework

1. **Engage a cross-functional group:** partners, IT, KM, and risk.

2. **Score each criterion** based on both product capabilities and vendor behavior.

3. **Document evidence** (e.g., citations, test drafts, integration testing) for transparency.

4. **Select solutions that meet professional standards first**, then evaluate efficiency gains.