Privilege Protection Setup Guide

The technical and operational controls that keep privileged client material from leaving your firm. Eight protocols, three deployment options, and a one-page test for whether your AI setup can handle privileged work.

From HitLai Institute — Module L2: "AI Workflows for Legal Operations"

WHAT'S ACTUALLY AT RISK

When privileged material is sent to a third-party AI service:

1. **Confidentiality breach** — Model Rule 1.6 violation, potentially sanctionable.
2. **Privilege waiver** — opposing counsel may argue privilege was waived by disclosure to a third party. Recent caselaw is unsettled but trending unfavorable.
3. **Attribution risk** — output of third-party AI may be discoverable in some jurisdictions.
4. **Contractual exposure** — most firm-client engagement letters include confidentiality undertakings that don't contemplate AI processors.

Translation: getting this wrong is catastrophic. Getting it right is not difficult — but it requires deliberate choices.

THE THREE DEPLOYMENT OPTIONS

For a firm using AI on privileged work, you have three viable architectures:

Option 1: Self-hosted on firm infrastructure (highest assurance)

- AI service runs on a server inside the firm's network.
- Local AI models via Ollama or vLLM — Llama, Mistral, Qwen, etc.
- Zero outbound API calls to any third party.
- Data never leaves the firm's network.

Pros: Maximum privilege protection, complete control, no third-party data-handling exposure. **Cons:** Requires IT capability and a server. Higher upfront cost. Local models are highly capable but typically a half-step behind frontier cloud models. **Best for:** Litigation firms, any firm with sensitive privileged matters, firms that need air-gapped deployments.

Option 2: Private-tenant cloud (high assurance)

- AI runs on a dedicated tenant in a managed cloud — your firm's data isolated from other tenants.
- Major AI providers' enterprise tiers (Anthropic, OpenAI, Google Vertex) with no-training contracts and data-residency commitments.
- Encrypted in transit and at rest. Audit logs available.

Pros: Frontier model quality. Lower IT burden. Acceptable for most non-extreme matter types. **Cons:** Data leaves the firm's network, requires careful contract review and ongoing vendor risk management. **Best for:** Mid-size firms doing transactional and regulatory work, firms whose matter mix doesn't include high-profile litigation.

Option 3: Hybrid (most realistic)

- Self-hosted for **privilege-tier work** (active litigation, sensitive client matters, opt-out clients).
- Private-tenant cloud for **non-privileged work** (legal research on public databases, marketing content, internal docs).
- Routing logic determines which pathway each matter uses.

Pros: Right tool for each matter. Cost-balanced. Pragmatic. **Cons:** Requires clear routing rules and training. Easy to mess up if the routing fails. **Best for:** Most firms over 25 attorneys.

Public AI services (free ChatGPT, free Claude, free Gemini) are NOT a viable option for any privileged work. They are first-party data processors with no contractual privilege protection. Their use on privileged material is an ethics-rule violation.

THE EIGHT PROTOCOLS

For any setup to be privilege-safe, you need eight controls in place:

1. Tool-approval list

What: A documented list of which AI tools are approved for which kinds of work — and which are explicitly banned. **Owner:** Managing partner or general counsel. **Update cadence:** Quarterly, or whenever a new tool is requested. **Test:** Can every attorney name the firm's approved tools? If no, the list isn't operational yet.

---

2. Anonymization protocol for any non-self-hosted use

What: Before any data flows to a non-self-hosted AI tool, replace client names, opposing parties, case numbers, and identifying facts with placeholders. **Owner:** Each attorney/paralegal using the tool, with training and a checklist. **Test:** Run a spot-check on 5 recent AI prompts — do they contain any identifiable client information?

---

3. No-training contracts with all third-party AI providers

What: Every contracted AI provider commits in writing that firm data is not used to train models. **Owner:** General counsel; reviewed annually. **Test:** Can you produce the relevant contract clause within 24 hours if asked?

---

4. Self-hosted deployment for privilege-tier matters

What: Active litigation, opt-out client matters, and any matter the partner classifies as sensitive run on self-hosted AI only. **Owner:** Practice-group leaders flag matters for the privilege tier; IT enforces the routing. **Test:** Can you prove no privilege-tier matter has had data leave the firm's network in the last 90 days? (Audit logs.)

---

5. Audit log for every AI interaction

What: Every AI prompt, every AI response, every approval action — logged with attorney name, matter ID, timestamp. **Owner:** IT operations; reviewed by managing partner monthly. **Test:** If a state bar inquiry asks "what did your AI do on Matter X on March 14?" — can you answer in 60 seconds?

---

6. Output verification before external use

What: Every AI output that leaves the firm — to a client, to a court, to opposing counsel — is verified by a qualified attorney first. Citations checked at the source. **Owner:** The reviewing attorney on each matter; partner-level supervision. **Test:** Has any AI-drafted output gone external without attorney verification in the last 60 days? (If yes, that's a sanctions risk.)

---

7. Opt-out routing

What: When a client opts out of AI assistance, their matter is flagged firm-wide and AI tools are blocked from running on their data. **Owner:** Intake; enforcement by IT. **Test:** Can you list every opt-out client and confirm no AI work has run on their matters?

---

8. Training and acknowledgment

What: Every attorney and paralegal completes AI-use training and signs an acknowledgment of the firm's AI policy before being granted AI tool access. **Owner:** Practice management; tracked annually. **Test:** Do you have a current acknowledgment on file for every active attorney and paralegal?

THE ONE-PAGE TEST

Use this to evaluate any proposed AI setup before approving it for privilege-tier work:

Question	Required answer
Does data leave the firm's network?	Self-hosted: No. Private-tenant: With contractual no-training and tenant-isolation. Public: Never.
Are AI prompts logged with attorney + matter attribution?	Yes
Is every external output reviewed by a licensed attorney before use?	Yes
Are citations verified at the source for every external use?	Yes
Is there a firm-wide tool-approval list, updated quarterly?	Yes
Is every attorney trained and acknowledged annually?	Yes
Can opt-out clients be flagged and routed AI-free?	Yes
Is privileged data anonymized before any non-self-hosted use?	Yes (or N/A if self-hosted only)
Does the audit log answer "what happened on this matter in the last 90 days?" in under 60 seconds?	Yes

If any answer is "No" or "I think so" — the setup is not yet privilege-safe. Fix the gap before deploying.

COMMON MISTAKES (LEARNED THE HARD WAY BY OTHER FIRMS)

❌ **"We just told the team to use ChatGPT carefully."** Carefully is not a control. Documented protocols are.

❌ **"The vendor said our data is safe."** Verbal assurances are not contractual. Get the no-training clause in writing.

❌ **"We anonymize most of the time."** "Most" is the gap. Either you have a protocol or you don't.

❌ **"We'll figure out logging later."** Without an audit trail, you cannot defend the firm against ethics inquiries. Build it before you deploy AI on privileged work.

❌ **"Our IT department says we don't need self-hosted."** This is a partner-level decision based on matter mix, not an IT decision based on infrastructure preferences. If your matter mix includes any litigation or any high-sensitivity client, evaluate self-hosted seriously.

❌ **"We've never had a problem."** Survivorship bias. The firms that have problems usually find out from a sanctions order or panel disqualification, not a near-miss.

WHAT GOOD LOOKS LIKE — AT 18 MONTHS IN

A firm running these protocols at the 18-month mark has:

- 3-5 active AI-assisted flows from the L2 catalog
- Self-hosted AI handling all litigation work; private-tenant cloud handling research and transactional starters
- Every attorney trained and acknowledged within the last year
- A monthly partner review of the AI audit log (5 minutes, not 5 hours)
- Engagement letters with the L1-02 disclosure clause as default
- Zero panel-disqualification risk on AI grounds
- A measurable productivity gain — typically 10-20% on the activities AI is genuinely good at, with no compromise of supervision or verification quality

This is achievable. It is not theoretical. It is a 90-day setup project for a small-to-mid firm.

THE HITLAI / AICTRLNET ANSWER, IN ONE LINE

> AI tools, your team, your systems — running together, safely.

For a law firm doing privilege-tier work: self-hosted deployment inside your firm's network, local AI via Ollama or vLLM, air-gapped if your matter requires. Privileged material never leaves your network. Every prompt logged. Every output verified by a qualified attorney before external use. Your DMS, billing, conflict, and engagement-letter systems all reflect AI use. Every step auditable for state-bar inquiries, partner reviews, or panel/RFP defenses.

Want a partner-meeting-ready walkthrough of THIS guide for your firm's matter mix? That's Module L2 — the 120-minute live setup session. → **hitlai.net/institute**

This guide is not legal advice. It is operational guidance based on ABA Formal Opinion 512 and current state-bar practice and is intended for firm-internal planning under the supervision of general counsel or ethics counsel.