Pleadly.

pleadly.ai

Architecture & Compliance Attestation

ABA Formal Opinion 512 (July 2024) — Generative Artificial Intelligence Tools
California COPRAC Practical Guidance — Generative AI
HIPAA Business Associate Agreement — Available on Request

Version	**1.3 — March 2026**
Document type	**Vendor Compliance Attestation**
Issued by	**Miko Labs LLC**
Operating as	**Pleadly.ai**
Effective	**March 2026**
Contact	**legal@mikolabs.ai**

CONFIDENTIAL — FOR LAW FIRM DUE DILIGENCE USE ONLY

This document does not constitute legal advice and is not a Business Associate Agreement.

1. EXECUTIVE SUMMARY

This document constitutes the formal Architecture and Compliance Attestation of Miko Labs LLC, operating as Pleadly.ai, issued for law firm vendor due diligence purposes in connection with ABA Formal Opinion 512 (July 2024), California COPRAC Practical Guidance on Generative Artificial Intelligence, and HIPAA obligations arising from the processing of protected health information in the form of client medical records.

Pleadly.ai is a managed AI demand letter intelligence platform for plaintiff-side personal injury law practices. The platform's architecture is designed from first principles around attorney-client privilege protection: all AI inference on privileged case content occurs exclusively on on-premises hardware operated by Miko Labs LLC. No privileged client information — including client names, injury descriptions, medical records, accident facts, police reports, or demand letter content — is transmitted to any third-party AI model provider, cloud inference service, or external API at any stage of processing.

This attestation is provided to assist law firms in conducting vendor due diligence. It is a factual description of architecture and operational practices as of the effective date. It does not constitute legal advice, and it is not a Business Associate Agreement (BAA). Firms requiring a BAA for HIPAA compliance should contact legal@mikolabs.ai. Firms should consult their own ethics counsel regarding jurisdiction-specific compliance obligations.

> **United States v. Heppner, No. 25-cr-00503-JSR (S.D.N.Y. Feb. 17, 2026) — Why Architecture Matters**
>
> On February 10, 2026, Judge Jed Rakoff of the U.S. District Court for the Southern District of New York ruled that 31 documents a criminal defendant generated using the consumer version of Claude — and later shared with defense counsel — were not protected by attorney-client privilege or the work product doctrine. Judge Rakoff issued a written memorandum opinion on February 17, 2026. Privilege failed not because the documents were AI-generated, but because the defendant used a public AI tool whose terms of service allow the provider to collect prompts and outputs, use them for model training, and disclose them to third parties including government regulatory authorities. Sharing privileged information under those terms constituted a third-party disclosure that defeated confidentiality under settled doctrine. Judge Rakoff noted the outcome might have differed had counsel directed the AI use in a manner analogous to a Kovel arrangement. The significance of this ruling for PI law practices using cloud AI tools is that the standard commercial terms of service of every major AI provider — OpenAI, Anthropic consumer, Google Gemini — include provisions that would trigger the same analysis. Pleadly's local inference architecture eliminates this risk at the infrastructure layer.

2. SCOPE AND PURPOSE

This attestation covers:

- The technical architecture of Pleadly.ai as of March 2026
- Data handling practices applicable to privileged client case content
- Security controls governing case data at rest and in transit
- The platform's relationship to ABA Formal Opinion 512, California COPRAC guidance, and HIPAA
- Firm responsibilities that remain with the engaging law firm

This attestation does not cover:

- The legal sufficiency of AI-generated demand letters — attorney review is required before any demand is transmitted
- Jurisdiction-specific ethics rules that may impose obligations beyond ABA Model Rules

— Conflict-of-interest screening — firms retain this responsibility entirely

— Court filing requirements regarding AI-generated content — firms must comply with local rules

3. SYSTEM ARCHITECTURE

Pleadly.ai is built on a three-plane architecture that strictly separates privilege-sensitive operations from non-privileged infrastructure. This separation is enforced at the network, hardware, and software layers — not only at the policy layer.

3.1 Control Plane

Hosted on: Vercel (Next.js application server) + Supabase (relational database)

The Control Plane handles authentication, firm administration, billing, case status tracking, and the attorney-facing dashboard. It processes only non-privileged metadata: user sessions, organization identifiers, case status flags, and billing relationships. The Control Plane never handles case content, medical records, demand letter text, or any other privileged information. All attorney-client privileged content is hard-blocked from this plane at the application layer. No AI inference occurs on this plane.

3.2 Intelligence Plane — Privileged Content Boundary

Hosted on: On-premises hardware, Miko Labs LLC facility

The Intelligence Plane is where all AI inference occurs. This plane runs exclusively on dedicated on-premises hardware operated by Miko Labs LLC at its secure facility. The language model serving demand drafting and document analysis runs entirely on this hardware with no public internet exposure. The Intelligence Plane is reachable only via authenticated internal network paths over a private VPN mesh. All AI inference on privileged content — document classification, text extraction, treatment gap analysis, liability assessment, verdict comparable matching, demand letter generation — occurs exclusively on this plane.

No case content is transmitted to OpenAI, Anthropic, Google, or any other third-party AI provider at any point in the processing pipeline.

3.3 Data Plane

Hosted on: Encrypted local storage, Miko Labs LLC facility

The Data Plane stores case files, demand drafts, firm-specific memory, and extracted case analytics. All storage volumes are protected by LUKS full-disk encryption. The Data Plane has no external replication. Case data does not transit any cloud storage provider.

NOTE: The critical architectural boundary: privileged case content — client names, injury details, medical records, police reports, demand letter text — never transits any external API, cloud service, or third-party server. This is enforced at the infrastructure layer. It is not solely a contractual commitment or a policy position.

4. TECHNICAL SAFEGUARDS

4.1 On-Premises AI Inference

All AI model inference for case-related tasks is performed on dedicated on-premises hardware operated by Miko Labs LLC. Current hardware specification: GMKtec EVO-X2, AMD Ryzen AI Max+ 395, 96GB unified memory. Primary reasoning model: Qwen3.5-35B-A3B, served via local inference server on a private non-public port. Embedding model: nomic-embed-text. No case content is transmitted to OpenAI, Anthropic, Google, or any third-party AI provider at any point in the processing pipeline.

4.2 HMAC-Signed Request Authentication

Every request to the Intelligence Plane is authenticated via HMAC-SHA256 signature using a shared secret provisioned at firm onboarding and rotatable on request. Requests without a valid HMAC signature are rejected before any AI processing occurs. Signature validation includes a 5-minute timestamp window to prevent replay attacks. The signature format is HMAC-SHA256("timestamp.body"), transmitted via X-Pleadly-Timestamp and X-Pleadly-Signature headers. Inbound callbacks from the Intelligence Plane to the Control Plane are also HMAC-verified and schema-validated before any database writes occur.

4.3 Encryption at Rest

All case data stored on Pleadly infrastructure is protected by LUKS (Linux Unified Key Setup) full-disk encryption. Logical volumes containing case files, demand drafts, and firm-specific memory are encrypted as children of an encrypted volume group. Encryption keys are never stored on the same physical media as the encrypted data. Standard backup media follows a 30-day purge cycle.

4.4 Encrypted Transit

All data in transit between client systems and Pleadly infrastructure is encrypted via TLS 1.3. Internal service communication is isolated to a private Tailscale mesh VPN. The Intelligence Plane has no inbound access from the public internet. Public-facing asset hosting (non-case-content marketing assets) is served via a separate Cloudflare Tunnel that has no access to case data or the Intelligence Plane.

4.5 Evidence Provenance Architecture

Every claim in a Pleadly-generated demand letter is linked to its source document by identifier, page number, and character offset. The system requires the AI model to return a source_document_id for each citation in its output. Outputs that include citations without recognized source document identifiers are rejected by post-generation validation before delivery to the attorney. This pipeline-layer control significantly reduces the risk of unsupported citations reaching the attorney's review queue. It does not eliminate the possibility that AI-generated content contains inaccuracies — attorney review remains required.

4.6 Multi-Tenant Isolation

Each law firm's case content is stored in a dedicated, organization-scoped vector collection. No case content from Firm A is ever in the same query scope as Firm B. Row-level security policies enforce organization isolation at the database layer, in addition to application-layer controls. All Intelligence Plane queries include an explicit organization identifier filter.

4.7 No Model Training on Client Data

Pleadly does not use client case data to train, fine-tune, or update AI models without explicit written consent from the firm. The attorney review workflow generates anonymized, PII-stripped signals only when the firm has separately and affirmatively opted in writing. Opting out has no effect on service quality or pricing. The on-premises model does not incorporate client data from one firm's cases into responses delivered to any other firm — this is the self-learning risk addressed by ABA Opinion 512, and it does not apply to Pleadly's architecture.

4.8 Audit Logging

All case-related API requests are logged to a write-only audit volume with restricted filesystem permissions. Logs capture request metadata — timestamp, case identifier, organization identifier, endpoint, processing time, response status — but do not capture the text content of privileged communications, including medical

record summaries, demand letter text, or case strategy content. Logs are retained for 90 days and are available for firm review upon written request.

4.9 Physical Security

The on-premises hardware serving the Intelligence Plane and Data Plane is located at a Miko Labs LLC controlled facility. Physical access is restricted to authorized Miko Labs LLC personnel. The hardware is not co-located in a shared data center. The network perimeter is protected by a stateful firewall. Remote access to the hardware for administrative purposes is conducted exclusively via the private Tailscale mesh VPN and requires multi-factor authentication.

4.10 Breach Notification

In the event of a confirmed security breach affecting client case content, Miko Labs LLC will notify the affected firm in writing within 72 hours of internal confirmation. Notification will include: the nature of the incident, the data categories potentially affected, the remediation steps taken, and recommended firm-side actions. This commitment covers confirmed breaches affecting privileged case content. Security events that do not affect case content (e.g., failed authentication attempts on the Control Plane) are logged internally and are not subject to this notification commitment absent actual data exposure.

4.11 Data Retention and Deletion

Case data and demand drafts are retained for the duration of the firm's active subscription plus 30 days to allow for data export. Upon written firm request or subscription termination, all firm data — case files, demand drafts, firm-specific memory — is deleted from active Pleadly infrastructure within 14 days. Backup media containing firm data is purged within 30 days of the deletion request under the standard backup rotation cycle. Audit log metadata, which does not contain privileged content, is retained for 90 days from the date of generation and is then automatically purged.

4.12 Material Architecture Change Notification

Miko Labs LLC commits to providing existing clients with at least 30 days written notice before implementing any material change to the architecture described in this attestation that would affect the privilege protection analysis — including but not limited to: adding a cloud-hosted AI inference component, adding a new subprocessor with access to case content, or changing the data residency of case files. The updated attestation will be published to pleadly.ai/security. Firms will be notified at their registered email address.

5. ABA FORMAL OPINION 512 — COMPLIANCE ANALYSIS

ABA Formal Opinion 512 (July 2024) addresses generative AI use in legal practice under the Model Rules of Professional Conduct. The opinion identifies two primary risk areas relevant to PI firms: (1) whether transmitting client data to cloud AI systems triggers disclosure obligations under Rule 1.6, and (2) whether AI tools that self-learn from client data require informed client consent. The following table summarizes Pleadly's architectural position on each relevant compliance factor.

Compliance Factor	Pleadly.ai	Cloud AI Tools (typical)
Case data leaves firm network	No — processed entirely on-premises	Yes — transmitted to provider cloud
Third-party AI provider access to case content	No — no external AI provider in pipeline	Yes — OpenAI / Anthropic / Google
ABA Op. 512 self-learning disclosure trigger	Not triggered — on-premises model, opt-in only	Present — most cloud PI AI tools trigger informed consent requirement

Compliance Factor	Pleadly.ai	Cloud AI Tools (typical)
Model training on client data	No — explicit written opt-in required	Varies; often default opt-in
Attorney-directed AI use	Yes — attorney directs all analysis through formal engagement	No — paralegal/client-operated SaaS
Encryption at rest	LUKS full-disk encryption	Provider-managed; varies
Audit log available to firm	Yes — 90-day metadata, on written request	Rarely
Breach notification to firm	Within 72 hours of confirmed breach affecting case content	Varies; often contractual only
Data deletion on termination	Within 14 days of written request	Varies; not always permanent
BAA available (HIPAA)	Yes — available on request at legal@mikolabs.ai	Varies; often unavailable or limited
Material architecture change notification	Yes — 30 days written notice before material changes	Rarely committed
CA COPRAC compliance	Architecture satisfies COPRAC security & isolation requirements	Requires firm-level evaluation

5.1 Rule 1.6 — Confidentiality of Information

Rule 1.6 requires attorneys to make reasonable efforts to prevent unauthorized disclosure of information relating to a representation. ABA Opinion 512 applies this to AI tool selection. Because Pleadly processes all privileged case content on on-premises hardware with no third-party AI provider in the pipeline, use of Pleadly does not create the third-party disclosure risk the opinion addresses. The Heppner analysis — where a cloud provider's TOS permitting data collection and disclosure to regulators defeated the confidentiality element of privilege — does not apply to Pleadly's architecture as described in this attestation.

5.2 Rule 1.1 — Competence

Rule 1.1 requires attorneys to maintain competence in technologies they use in practice. ABA Opinion 512 extends this to AI tools. This attestation, together with the architecture description at pleadly.ai/security, is provided to enable attorneys to evaluate the platform in light of their jurisdiction's ethics guidance prior to use.

5.3 Rules 5.1 and 5.3 — Supervisory Obligations

All Pleadly-generated demand letters are presented as drafts in the attorney review workflow. The platform is designed to deliver outputs to attorney review — not directly to opposing counsel — and requires affirmative attorney disposition (approve, revise, or reject) before any demand letter proceeds. The platform does not transmit demand letters to any external party autonomously.

6. HIPAA AND PROTECTED HEALTH INFORMATION

Personal injury cases routinely involve medical records, billing records, and treatment histories that may constitute protected health information (PHI) under the Health Insurance Portability and Accountability Act (HIPAA). Law firms handling PI cases are typically not covered entities under HIPAA, but may act as

business associates of covered entities (such as healthcare providers or insurers) in certain circumstances. Firms should consult with their own counsel regarding their HIPAA obligations.

6.1 Processing of Medical Records

Pleadly processes medical records uploaded by the firm as part of demand letter preparation. All such processing occurs on on-premises infrastructure as described in this attestation. Medical record content is not transmitted to any third-party AI provider or cloud inference service. Medical records are stored in encrypted form on the Data Plane and are subject to the same retention and deletion commitments described in Section 4.11.

6.2 Business Associate Agreement

Miko Labs LLC will execute a Business Associate Agreement (BAA) with any law firm that requires one as part of its HIPAA compliance program. Firms should contact legal@mikolabs.ai to request a BAA prior to using the platform with PHI. This attestation document is not a BAA and does not create BAA obligations.

> **NOTE:** This document is not a Business Associate Agreement. If your firm's HIPAA compliance program requires a signed BAA with AI vendors that process PHI, contact legal@mikolabs.ai before uploading medical records.

7. SUBPROCESSORS

Pleadly.ai uses a limited set of third-party subprocessors for non-privileged functions only. No subprocessor has access to case content, medical records, demand letter text, or any other privileged client information.

Subprocessor	Function	Data Accessed (no case content)
Vercel	Control Plane application hosting	Authentication tokens, session management. No case content.
Supabase	Control Plane database	User accounts, firm metadata, billing relationships. No case content.
Stripe	Subscription billing	Payment and subscription data. No case content.
Resend	Transactional email (status notifications)	Notification type and case identifier (no case content, no PHI). Subject lines contain only status codes.

All AI inference, case data storage, and demand generation occur exclusively on Miko Labs LLC on-premises infrastructure, with no subprocessor involvement.

8. FIRM RESPONSIBILITIES

Pleadly's architecture addresses the primary third-party disclosure risk identified in ABA Opinion 512. The following responsibilities remain with the engaging law firm and are not affected by Pleadly's architecture.

8.1 Supervisory Review of AI Outputs (Rules 1.1, 5.1, 5.3)

All Pleadly-generated demand letters are drafts. They must be reviewed by a licensed attorney before transmission to opposing parties, insurers, or any third party. AI-generated content may contain inaccuracies despite evidence provenance controls. The attorney is responsible for verifying all factual assertions, damages calculations, citations, and legal standards before approving any demand.

8.2 Client Consent and Disclosure (Rules 1.4, 1.6)

Attorneys should evaluate whether disclosure of AI use is required under their engagement agreements, applicable state bar rules, or because AI output will influence a significant decision in a representation. Pleadly's architecture does not trigger the self-learning informed consent requirement of ABA Opinion 512. However, jurisdictional disclosure rules vary, and firms should consult their ethics counsel.

8.3 Conflict of Interest Checks

Pleadly does not perform conflict-of-interest checks. Firms must maintain their own conflict-check and matter-opening procedures before adding a new matter to the platform.

8.4 Jurisdiction-Specific Ethics Rules

Firms must comply with any state bar ethics rules regarding AI use, client disclosure, or court filing requirements that differ from or exceed ABA Model Rules. California attorneys should review COPRAC Practical Guidance on Generative AI alongside this attestation. Attorneys in other jurisdictions should consult their applicable bar guidance.

8.5 Court Filing Requirements

Some jurisdictions and individual courts have adopted rules requiring disclosure of AI use in court filings, certification that AI-generated content has been reviewed for accuracy, or specific formatting requirements for AI-assisted documents. Firms are responsible for compliance with all applicable local rules. Pleadly outputs used in court filings must be reviewed and certified by the filing attorney.

8.6 HIPAA (if applicable)

Firms that are covered entities or business associates under HIPAA and that upload PHI to Pleadly should execute a BAA with Miko Labs LLC prior to use. See Section 6.

9. FORMAL ATTESTATION

FORMAL ATTESTATION — Miko Labs LLC

Miko Labs LLC hereby attests that, as of the effective date of this document:

1. No privileged client case content processed through Pleadly.ai is transmitted to any third-party AI model provider, cloud inference service, or external API.
2. All AI inference on case-related content occurs exclusively on on-premises hardware operated by Miko Labs LLC, isolated from the public internet.
3. Case data is encrypted at rest using LUKS full-disk encryption and in transit using TLS 1.3.
4. Pleadly does not use client case data to train or fine-tune AI models without explicit written consent from the firm.
5. Audit logs of all case-related processing are maintained for 90 days and are available to the firm upon written request. Logs do not contain the text of privileged communications.

6. In the event of a confirmed security breach affecting client case content, Miko Labs LLC will notify the affected firm in writing within 72 hours of internal confirmation.

7. Upon subscription termination or written firm request, all firm case data will be deleted from active Pleadly infrastructure within 14 days, and from backup media within 30 days under the standard backup rotation cycle.

8. Each law firm's case content is stored in dedicated, organization-scoped storage and query collections. No case content from one firm is accessible to any other firm.

9. Miko Labs LLC will provide at least 30 days written notice to existing clients before implementing any material change to the architecture described in this attestation that would affect the privilege protection analysis.

10. Miko Labs LLC will execute a Business Associate Agreement with any law firm that requires one for HIPAA compliance purposes. This attestation is not itself a BAA.

Miko Labs LLC
Operating as Pleadly.ai
Effective: March 2026 | Version: 1.3 | Contact: legal@mikolabs.ai

10. DISCLAIMER AND LIMITATIONS

This attestation reflects the technical architecture and operational practices of Pleadly.ai as of the effective date stated on the cover page. It is provided for informational purposes to assist law firms in conducting vendor due diligence under ABA Formal Opinion 512, California COPRAC Practical Guidance on Generative AI, applicable state bar guidance, and HIPAA.

This document does not constitute legal advice. It does not represent that use of Pleadly.ai will satisfy any specific legal or regulatory obligation of the engaging firm. Firms should consult with their own ethics counsel and, where applicable, HIPAA counsel regarding their compliance obligations.

This document is not a Business Associate Agreement. It does not create BAA obligations on either party. Firms requiring a BAA should request one separately.

The legal and regulatory landscape governing AI use in legal practice is evolving. Miko Labs LLC monitors relevant developments, including ABA ethics opinions, state bar guidance, and court rules, and will update this attestation when material changes occur. The then-current version of this attestation is available at pleadly.ai/security. Existing clients will be notified of material updates at their registered email address.

AI-generated outputs from Pleadly.ai may contain inaccuracies. Evidence provenance controls reduce but do not eliminate this risk. Attorney review of all AI-generated content before use or transmission is required.

Effective: March 2026 | Version: 1.3 | pleadly.ai/security | Questions: legal@mikolabs.ai