Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: 7 October 2026

T. Adebayo
O. Apalowo
F. Makanjuola
Veridom Ltd
5 April 2026

OMP Domain Profile: Legal AI Supervision Under ABA Model Rule 5.3 and California Senate Bill 574
draft-veridom-omp-legal-00

Abstract

This document defines a domain profile of the Operating Model Protocol (OMP) for legal AI deployments subject to attorney supervision obligations under ABA Model Rule 5.3 Responsibilities Regarding Nonlawyer Assistance and California Senate Bill 574 (SB 574, effective January 1, 2026). These instruments impose principal accountability requirements on attorneys who use AI tools to assist with legal work product -- requiring attorneys to verify AI-generated material, ensure compliance with professional duties, and maintain evidence of supervision.

This profile specifies how OMP's deterministic routing invariant, Watchtower enforcement framework, and three-layer cryptographic integrity architecture satisfy the attorney supervision obligations imposed by Rule 5.3 and SB 574, and defines the domain-specific Watchtower configurations, Named Accountable Officer assignments, and Audit Trace schema extensions applicable to legal AI deployments. The profile is designated the CiteGuard profile.

The OMP core specification is defined in a separate Internet-Draft.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 7 October 2026.

Table of Contents

1. Introduction

The deployment of AI in legal practice has accelerated substantially since 2024, driven by improvements in large language model capabilities for legal research, contract analysis, brief drafting, and citation generation. Law firms, corporate legal departments, and legal technology companies now routinely use AI systems to assist with work product that bears attorney signatures and carries professional and legal accountability.

Two instruments have crystallised the attorney supervision obligations that apply to AI-assisted legal work:

* ABA Model Rule 5.3, [ABA-RULE-5-3] as clarified by ABA Formal Opinion 512 [ABA-OP-512] (July 2023), establishes that attorneys must make reasonable efforts to ensure that the conduct of nonlawyer assistance -- including AI tools -- is compatible with the attorney's professional obligations. Attorneys must understand the capabilities and limitations of AI tools, verify AI-generated material for accuracy, and maintain a supervisory relationship over AI outputs that become part of legal work product.

* California Senate Bill 574 (effective January 1, 2026) extends these obligations with specific requirements: attorneys must ensure that confidential client information is not disclosed to public AI systems, must verify and correct AI-generated material, must remove AI-generated content that may contain bias, and must personally verify citations and case references included in submitted filings. SB 574 also prohibits arbitrators from delegating decision-making authority to AI systems.

These instruments impose a structural evidence requirement: an attorney who relies on AI assistance for legal work product must be able to demonstrate, if challenged, that they supervised the AI tool, reviewed its output, exercised independent professional judgment, and corrected errors before the work product was submitted or delivered.

The Operating Model Protocol (OMP) [I-D.veridom-omp] is a deterministic decision-enforcement protocol that generates a tamper-evident Audit Trace at the point of every AI-assisted decision. Applied to legal AI deployments, OMP provides the evidence infrastructure that makes attorney supervision provable rather than merely asserted.

This document defines the CiteGuard profile: the domain-specific instantiation of OMP for legal AI supervision under Rule 5.3 and SB 574. The name reflects the profile's primary enforcement focus:

ensuring that every citation, reference, and claim in AI-assisted legal work product is verifiably reviewed by a named supervising attorney before delivery or filing.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119] [RFC8174].

2. Terminology

This document uses the terminology defined in [I-D.veridom-omp]. In addition:

Supervising Attorney The licensed attorney who bears professional responsibility for an AI-assisted legal interaction under Rule 5.3 or SB 574. In OMP terms, the Supervising Attorney is the Named Accountable Officer for ASSISTED and ESCALATED interactions.

Legal Work Product Any document, analysis, draft, filing, research output, or communication produced with AI assistance that is delivered to a client, submitted to a court or arbitral tribunal, or used in a legal proceeding.

AI-Assisted Legal Interaction Any interaction in which an AI system contributes to the generation, verification, analysis, or citation of legal content that becomes or may become part of Legal Work Product.

Citation Verification The act of a Supervising Attorney confirming that a case citation, statutory reference, regulatory citation, or other legal authority cited in AI-generated content accurately represents the cited source and is applicable to the stated proposition.

CiteGuard Invariant The two-property invariant defined in Section 5: every AI-assisted legal interaction is routed to ASSISTED or ESCALATED (never AUTONOMOUS for Legal Work Product), and every routing produces a sealed, independently verifiable CiteGuard Audit Trace.

Privilege Review Flag A field in the CiteGuard Audit Trace indicating whether the interaction involved content subject to attorney-client privilege or work product doctrine.

3. Legal Framework Analysis

3.1. ABA Model Rule 5.3

ABA Model Rule 5.3 requires that attorneys with supervisory authority over nonlawyer assistants make reasonable efforts to ensure that the assistants' conduct is compatible with the professional obligations of the attorney. ABA Formal Opinion 512 (July 2023) applies this obligation to AI tools used in legal practice.

The supervision obligation under Rule 5.3 has three components relevant to AI deployments:

* *Competence obligation:* An attorney who uses an AI tool must understand the tool's capabilities and limitations to a degree sufficient to supervise its outputs. This includes understanding the tool's propensity to generate hallucinated citations, its training data cutoff, and its limitations with jurisdiction-specific law.

* *Verification obligation:* An attorney must review AI-generated work product for accuracy before delivery or use. For citations, this means personally verifying that the cited authority exists and supports the stated proposition.

* *Accountability obligation:* The attorney bears professional responsibility for AI-generated work product delivered under the attorney's name. The attorney cannot delegate this accountability to the AI tool.

3.2. California SB 574

California Senate Bill 574, [CA-SB574] effective January 1, 2026, imposes attorney supervision requirements specific to California practice:

* *Confidentiality:* Attorneys must ensure that confidential client information is not entered into public AI systems without client consent.

* *Verification and correction:* Attorneys must personally verify AI-generated material and correct any errors, inaccuracies, or misleading content.

* *Bias removal:* Attorneys must review AI-generated content for potential bias and remove any biased analysis.

* *Citation verification:* Attorneys must personally verify all citations, case references, and statutory references included in filings or documents submitted to courts or arbitral tribunals.

* *Non-delegation:* Arbitrators and attorneys acting as decision-makers may not delegate the decision itself to an AI system.

3.3. Convergent Requirements

Rule 5.3 and SB 574, taken together, define a structure that maps directly onto OMP's three routing states:

* AI-generated content reviewed and approved by the Supervising Attorney without modification corresponds to the ASSISTED routing state.

* AI-generated content requiring correction, subsequently corrected and approved, corresponds to the ESCALATED routing state with resolution.

* Any AI-generated interaction where supervision evidence cannot be produced corresponds to the structural gap that OMP closes.

Under this profile, there are no AUTONOMOUS routing outcomes for AI-Assisted Legal Work Product interactions. Every such interaction MUST be routed to ASSISTED or ESCALATED.

4. OMP CiteGuard Profile

4.1. Routing States Under This Profile

AUTONOMOUS NOT PERMITTED for AI-Assisted Legal Interactions under this profile. Implementations MUST configure WT-LEGAL-01 as a universal FORCE_ASSISTED trigger for all interactions classified as Legal Work Product. AUTONOMOUS routing is reserved for non-Legal-Work-Product interactions only.

ASSISTED The standard routing state for AI-Assisted Legal Interactions. The Supervising Attorney's identity, review timestamp, approval decision, and any corrections are recorded in the CiteGuard Audit Trace.

ESCALATED Triggered by Watchtower detection of a confidentiality breach, citation verification failure, hallucinated authority, detected bias, or non-delegation violation. The AI system's output MUST NOT be delivered or filed until the Supervising Attorney has reviewed, corrected, and approved.

4.2. Named Accountable Officer: The Supervising Attorney

Under the CiteGuard profile, the Named Accountable Officer for every ASSISTED and ESCALATED interaction is the Supervising Attorney. The Supervising Attorney MUST be a licensed attorney in the jurisdiction where the Legal Work Product will be used.

The following fields are REQUIRED in the Supervising Attorney record:

* supervising_attorney_id: unique deployment identifier;
* supervising_attorney_bar_jurisdiction: ISO 3166-2 codes for licensed jurisdictions;
* review_timestamp: ISO 8601 UTC of the review action;
* review_decision: one of APPROVED, APPROVED_WITH_CORRECTIONS, RETURNED_FOR_REWORK;
* corrections_summary: REQUIRED if review_decision is not APPROVED.

4.3. Watchtower Definitions

4.3.1. WT-LEGAL-01: Supervising Attorney Gate

Trigger: Any AI-Assisted Legal Interaction.

Action: FORCE_ASSISTED.

Rationale: Rule 5.3 and SB 574 impose non-waivable attorney supervision obligations. This Watchtower makes it architecturally impossible for an AI-Assisted Legal Interaction to proceed to delivery or filing without generating a supervision evidence record. It cannot be disabled for Legal Work Product interactions.

4.3.2. WT-LEGAL-02: Confidentiality Boundary Gate

Trigger: Interaction payload contains client confidential information destined for an AI system outside an approved confidentiality boundary.

Action: HARD_BLOCK.

Rationale: SB 574 requires attorneys to ensure that confidential client information is not entered into public AI systems. Rule 1.6 applies independently. HARD_BLOCK ensures violations cannot occur without a blocking record.

4.3.3. WT-LEGAL-03: Citation Verification Gate

Trigger: Output contains citations to legal authorities not yet verified against an accessible source within the current interaction.

Action: FORCE_ESCALATED.

Rationale: SB 574 requires attorneys to personally verify citations in filings. This Watchtower enforces that obligation structurally: AI-generated content with unverified citations cannot be approved without an attorney citation verification record.

4.3.4. WT-LEGAL-04: Hallucination Detection Gate

Trigger: Output contains a citation, case name, or legal authority that cannot be located in accessible legal databases, or where the cited passage does not appear at the cited location.

Action: HARD_BLOCK for submissions; FORCE_ESCALATED for drafts.

Rationale: AI hallucination of legal citations is a documented pattern resulting in court sanctions and professional discipline. This Watchtower provides pre-submission enforcement. The CiteGuard Audit Trace records the unverifiable citation, the database query result, and the Supervising Attorney’s disposition.

4.3.5. WT-LEGAL-05: Bias Detection Gate

Trigger: Operator’s bias detection module flags potential biased analysis, discriminatory framing, or stereotyped characterisation.

Action: FORCE_ESCALATED.

Rationale: SB 574 requires attorneys to remove AI-generated content that reflects bias. The Watchtower ensures bias flags generate a supervision record with Supervising Attorney disposition.

4.3.6. WT-LEGAL-06: Non-Delegation Gate

Trigger: AI output constitutes or is intended to constitute a final decision in a matter where an attorney or arbitrator is the designated decision-maker (arbitral award, legal opinion delivered as final determination).

Action: HARD_BLOCK.

Rationale: SB 574 prohibits arbitrators from delegating decision-making authority to AI systems. Rule 5.3 requires independent professional judgment. The AI system's analysis may inform the decision as ASSISTED input, but the decision record MUST reflect the human decision-maker's independent judgment.

4.4. Audit Trace Schema Extensions

The following fields are REQUIRED in the Audit Trace schema under the CiteGuard profile, in addition to the core fields defined in [I-D.veridom-omp] Section 7:

supervising_attorney_id string, REQUIRED for ASSISTED and ESCALATED outcomes.

supervising_attorney_bar_jurisdiction string, REQUIRED. Comma-separated ISO 3166-2 codes. Example: "US-CA,US-NY".

review_timestamp string, ISO 8601 UTC, REQUIRED for ASSISTED and ESCALATED outcomes.

review_decision string, REQUIRED. One of: APPROVED, APPROVED_WITH_CORRECTIONS, RETURNED_FOR_REWORK.

corrections_summary string, OPTIONAL if APPROVED; REQUIRED otherwise.

citations array of objects, REQUIRED if the interaction generated legal citations. Each object MUST contain: citation_text, source_verified (boolean), verification_method, verification_timestamp (ISO 8601 UTC), verified_by (one of: "AI_SYSTEM", "SUPERVISING_ATTORNEY").

work_product_type string, REQUIRED. RECOMMENDED values: "court_filing", "client_advice", "contract_draft", "legal_research", "arbitral_submission", "internal_memo".

privilege_review_flag boolean, REQUIRED. True if the interaction involved potentially privileged content.

confidentiality_boundary_verified boolean, REQUIRED. True if WT-LEGAL-02 evaluated the target AI system.

profile_version string, REQUIRED. MUST be "VERIDOM-CITEGUARD-v1.0" for this profile version.

5. The CiteGuard Invariant

Implementations of this profile MUST satisfy the following two-property invariant:

Property 1 (Supervision completeness) Every AI-Assisted Legal Interaction that contributes to Legal Work Product MUST generate a sealed CiteGuard Audit Trace containing a Supervising Attorney review record before the work product is delivered or filed.

Property 2 (Immutable trail) The CiteGuard Audit Trace MUST be sealed with the three-layer integrity architecture defined in [I-D.veridom-omp] Section 7 (SHA-256 chain, RFC 3161 TimeStampToken, institution signature). Any modification to any historical Audit Trace record MUST be detectable by any third party without access to the operator’s or implementer’s infrastructure.

These two properties mean that for any AI-Assisted Legal Interaction processed under this profile, an attorney facing a Rule 5.3 or SB 574 compliance inquiry can produce: (a) a sealed, tamper-evident record of the specific AI output; (b) the Supervising Attorney’s identity, review timestamp, and decision; (c) citation verification records for every citation in the output; (d) Watchtower evaluation results; and (e) an independently verifiable integrity proof that the records have not been modified since sealing.

6. Proof-Point as Supervision Evidence

The OMP Proof-Point artefact generation mechanism (defined in [I-D.veridom-omp] Section 7.5) produces a self-contained supervision evidence package for any defined time window. Under this profile, the Proof-Point artefact for a legal deployment MUST include, for each AI-Assisted Legal Interaction: the full CiteGuard Audit Trace, the Supervising Attorney review record, citation verification records, Watchtower evaluation log, chain integrity proof (SHA-256 Merkle root), and RFC 3161 TimeStampToken verification output from the OMP Reference Validator [OMP-OPEN-CORE].

This artefact is designed to be self-contained: a disciplinary authority, court, or malpractice insurer with no access to the operator’s systems can verify its integrity and completeness using only the OMP Reference Validator and the public key material of the Timestamp Authority.

7. Interaction with Legal Privilege

CiteGuard Audit Trace records may contain information subject to attorney-client privilege or work product doctrine. Operators MUST apply the privilege_review_flag field. The existence of the Audit Trace does not waive privilege; the records were created as part of the supervisory process, not for disclosure to adverse parties.

The chain integrity proof (Merkle root and TimeStampToken) can be disclosed to demonstrate that a complete Audit Trace exists and has not been tampered with, without disclosing the content of individual records. This allows attorneys to assert the integrity of their supervision records without waiving privilege over their content.

8. Security Considerations

The security considerations of [I-D.veridom-omp] apply in full to this profile.

Supervising attorney identity: Operators MUST ensure that supervising_attorney_id values cannot be spoofed or assigned to non-attorneys within the deployment system.

Review timestamp integrity: The review_timestamp field MUST be set by the OMP pipeline at the time of the review action. Operators MUST ensure the pipeline clock is monotonic and cannot be manipulated to backdate supervision records.

Citation database availability: WT-LEGAL-03 and WT-LEGAL-04 depend on legal database access. Operators MUST treat database unavailability as a C_d reduction event, routing interactions to ESCALATED where citation verification cannot be performed.

9. IANA Considerations

This document has no IANA actions.

10. References

10.1. Normative References

[I-D.veridom-omp]
Adebayo, T., Apalowo, O., and F. Makanjuola, "Operating Model Protocol (OMP): A Deterministic Decision-Enforcement Protocol with Externalized Proof-of-Integrity", Work in Progress, Internet-Draft, draft-veridom-omp-00, March 2026, <https://datatracker.ietf.org/doc/html/draft-veridom-omp-00>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

10.2. Informative References

[ABA-OP-512] ABA Standing Committee on Ethics and Professional Responsibility, "Formal Opinion 512: Generative Artificial Intelligence Tools", July 2023.

[ABA-RULE-5-3] American Bar Association, "ABA Model Rules of Professional Conduct, Rule 5.3: Responsibilities Regarding Nonlawyer Assistance", 2023.

[CA-SB574] California Legislature, "Senate Bill 574: Attorneys: Artificial Intelligence", January 2026.

[I-D.veridom-omp-euaia] Adebayo, T., Apalowo, O., and F. Makanjuola, "OMP Domain Profile: EU AI Act Article 12 Logging and Traceability Requirements for High-Risk AI System Operators", Work in Progress, Internet-Draft, draft-veridom-omp-euaia-00, April 2026, <https://datatracker.ietf.org/doc/html/draft-veridom-omp-euaia-00>.

[OMP-OPEN-CORE] Veridom Ltd, "OMP Open Core: Reference Validator and Schema Library", Apache 2.0, https://github.com/veridomltd/omp-open-core, 2026.

[ZENODO-OMP] Adebayo, T., Apalowo, O., and F. Makanjuola, "OMP -- Operating Model Protocol: A Deterministic Routing Invariant for Tamper-Evident AI Decision Accountability in Regulated Industries", Zenodo DOI 10.5281/zenodo.19140948, March 2026.

Authors' Addresses

Tolulope Adebayo
Veridom Ltd
London
United Kingdom
Email: tolulope@veridom.io

Oluropo Apalowo
Veridom Ltd
Awka
Nigeria
Email: ropo@veridom.io

Festus Makanjuola
Veridom Ltd
Toronto
Canada
Email: festus@veridom.io