Frame Infrastructure for AI, Vol. 3: Expert Oversight Network — Traceable Human Judgment for AI-Native Professional Systems

Author: Xiangyu Guo – Independent Researcher – © 2026 Xiangyu Guo. Creative Commons Attribution Non Commercial No Derivatives 4.0 International

Executive Summary

This article develops the Expert Oversight Network (EON) as the professional verification layer of the Frame Infrastructure for AI series. Frame Theory defines trust as the product of Presence and Integrity. Vol. 1 argued that Presence, for machine systems, must take the form of source-anchored structural persistence rather than temporary context reconstruction. Vol. 2 argued that Integrity, for consequential AI systems, must take the form of traceable responsibility rather than symbolic safety statements. Vol. 3 adds the missing human layer: licensed, auditable, and risk-calibrated expert judgment integrated into machine-native memory and into a liability architecture that can support insurance, procurement, and regulated deployment. The original EON article introduced the paradigm of certified-expert auditing, transparent review labels, and feedback loops from expert correction into model improvement. This article converts that paradigm into an explicit system specification.

The problem is operational, not rhetorical. Professional AI systems now produce recommendations in medicine, law, finance, and public administration, but current deployments usually expose only the final output and a thin explanation surface. They often do not preserve a structured record of which claims were sourced, which claims were inferred, which expert verified or rejected which claim, what disagreements remained unresolved, which actor had override authority, and what final state entered the audit trail. NIST places accountability and transparency at the center of trustworthy AI, and explicitly treats role clarity, documentation, independent review, and ongoing monitoring as core risk-management practices. The EU AI Act similarly requires logging, transparency to deployers, and effective human oversight for high-risk

systems. Those sources establish the regulatory direction. They do not by themselves specify a deployable architecture for expert verification in professional domains. EON fills that gap.

EON should not be understood as “a human checks the answer sometimes.” It is a formal review system with five properties. First, expert review is bound to machine-native memory, so that verification attaches to claims, sources, updates, and conflicts rather than to a free-floating answer. Second, review is risk-calibrated: some outputs are sampled randomly; some are sampled by risk; some cannot be released without licensed human sign-off. Third, disagreement is preserved as a first-class object instead of being averaged away. Fourth, every review action writes to an audit trail suitable for incident reconstruction, post-market monitoring, and contractual evidence. Fifth, the review state is visible to downstream actors, including users, deployers, auditors, and insurers. Those properties align closely with W3C’s provenance model, which defines provenance as a record of the entities, activities, and agents involved in producing a thing and treats that record as central to trust judgments. They also align with NIST’s recent generative-AI guidance, which treats governance, content provenance, pre-deployment testing, and incident disclosure as primary considerations.

The central claim of this article is narrow and practical. High-risk professional AI becomes more governable when expert judgment is made traceable, stateful, and allocable. The result is not perfect safety. It is a system in which review status, override authority, failure location, and duty boundaries become legible enough to manage. That legibility is the minimum condition for disciplined deployment, contractual allocation, and insurability.

Position and Problem Statement

This article occupies a precise place in the larger framework. Frame Theory states the structural law: trustworthy AI requires Presence × Integrity. Vol. 1, *Beyond Human-Like Memory: Toward a Machine-Native Architecture of Structural Persistence* (DOI: 10.5281/zenodo.20738963), translated Presence into an architecture of persistent sources, states, claims, relations, and updates. Vol. 2, *From Black-Box Safety to Traceable Responsibility: Machine-Native Memory, Liability Allocation, and the Insurability of AI Risk* (DOI: 10.5281/zenodo.20738983), translated Integrity into a layered responsibility architecture in which sources, transformations, outputs,

and decision nodes can be located. Vol. 3 connects those two results to professional practice: it specifies how human judgment enters the system as a traceable verification event rather than as an informal fallback. The original EON article established the oversight paradigm; the present article defines its operational form.

The missing piece in current professional AI is not additional normative language. It is the absence of a verifiable judgment node. NIST's AI RMF states that trustworthy AI depends on accountability and that accountability presupposes transparency. It also treats documentation, independent review, clearly defined human roles, and the ability to track risks over time as operational requirements rather than optional reporting. The EU AI Act is directionally consistent: high-risk AI systems must support automatic logs over the system lifetime; those logs must provide traceability appropriate to the intended purpose; and high-risk systems must be designed so that natural persons can effectively oversee them, remain aware of automation bias, interpret outputs correctly, and disregard or interrupt system outputs when necessary.

That creates a concrete design problem. In a professional setting, an AI answer is rarely a single undifferentiated object. It contains sourced propositions, inferred propositions, jurisdiction-specific assumptions, temporary preferences, and sometimes unresolved ambiguity. If an expert later reviews the answer, the system must answer at least six questions: what exactly was reviewed; what evidence supported it; what was changed; whether disagreement remained; who had authority to release the final version; and what duties attached to that release state. Without that structure, "human oversight" becomes a presentation layer. It looks compliant, but it does not support post hoc reconstruction, incident accountability, or insurance underwriting.

The problem is more acute in professional advice because the relevant harm often depends on context and timing. A clinical suggestion may be acceptable as general education but unacceptable as patient-specific treatment advice. A legal summary may be useful as a background note but impermissible as a court filing without human verification. A financial recommendation may be informative at low value but not at a threshold where a regulated entity must justify suitability, validation, or board-level review. Public-sector AI may be safe for back-office triage but not for decisions that affect benefits, liberty, or public safety without

impact assessment, independent evaluation, and additional human oversight. Sectoral rules differ, but the architecture problem is shared.

The precise problem statement is therefore as follows: current AI deployments in professional domains lack a standardized mechanism for binding expert review to claim-level provenance, state transitions, conflict preservation, and liability allocation. EON is proposed as that mechanism.

System Specification

EON is a review layer that sits above generation and below consequential release. It is not a general ethics committee and not a post hoc complaint inbox. It is a structured service that receives AI outputs, classifies them by risk, routes some or all of them to verified human experts, records the experts' judgments as state-changing events in machine-native memory, and exposes review status to downstream actors. In architectural terms, EON is the human verification subsystem for the memory and responsibility layers specified in Vol. 1 and Vol. 2. It turns human judgment into a persistent, attributable object.

The minimal EON stack has seven components. The generation layer produces an initial output and claim graph. The memory layer stores source objects, claim objects, inference markers, prior review states, and update history. The risk-routing layer assigns a review mode based on domain, user context, actionability, monetary or rights threshold, uncertainty, and prior incident data. The credentialing layer determines which experts are eligible to review which outputs. The review layer presents the relevant materials to experts through a controlled interface. The synthesis layer resolves confirmation, correction, disagreement, and escalation states. The audit layer writes immutable or tamper-evident records for verification, override, incident response, and external inspection. This is consistent with NIST's repeated emphasis on lifecycle-wide documentation, role separation, TEVV, monitoring, and independent evaluation, and with W3C PROV's treatment of entities, activities, and agents as the core objects of provenance.

The core release workflow is straightforward. An AI output is first decomposed into claims linked to sources or marked as inferred. The system then assigns a risk tier. Low-risk outputs

may be released immediately and entered into a random or periodic sampling pool. Medium-risk outputs may be released with a provisional state and post-delivery review. High-risk outputs should be held until expert review is complete. Critical outputs—those that trigger a legal filing, a patient-specific treatment step, a material financial execution, or a public decision affecting rights or safety—should require explicit human sign-off before release or execution. The final user-visible object is never just "the answer." It is the answer plus a review state: unreviewed, sampled, reviewed-confirmed, reviewed-corrected, conflict-open, escalated, or signed off. That state should be machine-readable and contractually meaningful.

Expert review in EON is claim-directed, not merely answer-directed. The system should be able to ask a reviewer to confirm or correct a specific proposition, bracket a claim as unsupported, identify omitted considerations, or mark a recommendation as outside the reviewer's licensed scope. That matters because professional review is often partial. A lawyer may verify authority and jurisdictional fit while declining an economic assumption. A physician may verify triage logic while flagging that final diagnosis requires unavailable patient data. The system should therefore preserve review scope explicitly instead of overstating certainty. This approach is consistent with NIST's call for role-appropriate transparency, documented limits, and context-specific human-AI interaction, and with the AI Act's requirement that deployers receive the information necessary to interpret and use outputs appropriately.

EON also requires a clearer definition of outcome states than current "human in the loop" language usually provides. There are only four operationally meaningful endings for a review transaction. The output may be confirmed without substantive change. It may be corrected and rereleased as a superseding version. It may remain disputed and be escalated to a higher reviewer class or second panel. Or it may be blocked from release or execution because it exceeds system scope or fails an evidence threshold. Anything less precise collapses review into commentary rather than governance. OMB's federal AI memorandum is instructive on this point: agencies are required to complete impact assessments, perform real-world testing, obtain independent evaluation from a reviewing authority that was not directly involved in development, and add human oversight and fail-safes where rights or safety are materially affected. EON operationalizes a similar logic for professional services outside the public sector.

Data and Operating Rules

The decisive move in EON is to convert expert oversight into structured data rather than leaving it as prose or organizational memory. The data model below is the minimum needed for provenance, verification, conflict preservation, update history, and auditability.

Record type	Required fields	Primary writer	Function in EON
SourceRecord	`source_id`, source hash or URI, source type, jurisdiction, version, timestamp, access rights, reliability class	Ingestion service	Anchors claims to specific source material and versions
ClaimRecord	`claim_id`, `output_id`, text span, source links, `inference_flag`, confidence or uncertainty field, domain tags	Generation service	Separates sourced claims from inferred claims within one output
ReviewTask	`task_id`, `output_id`, risk tier, routing reason, required specialty, jurisdiction requirement, deadline, blinded review payload, review scope	EON coordinator	Specifies what is to be reviewed, by whom, and under what standard
VerificationRecord	`verification_id`, reviewer credential reference, verdict (`confirm`, `correct`, `reject`, `outside_scope`), rationale, edited text or annotations, timestamp	Expert interface	Stores the expert's judgment as a traceable event
ConflictRecord	`conflict_id`, linked verification records, materiality rating, disputed claims, escalation status, resolution rule	EON coordinator	Preserves expert disagreement rather than suppressing it

UpdateEvent	`update_id`, prior object reference, superseding object reference, diff hash, reason code, approval node, timestamp	System or expert reviewer	Maintains state history when an answer is changed
SignOffRecord	`signoff_id`, authorized natural person, license/jurisdiction match, threshold basis, final disposition, override rationale, timestamp	Licensed reviewer or deployer authority	Marks outputs that cannot be released or executed without explicit human authorization
AuditLogEntry	`event_id`, actor, action, affected object ids, session context, access context, immutable checksum, timestamp	Platform logging layer	Supports incident reconstruction, regulator requests, insurer review, and contractual evidence

This model is intentionally compatible with the abstractions in W3C PROV. A `SourceRecord` and `ClaimRecord` are entities; `ReviewTask`, `VerificationRecord`, `UpdateEvent`, and `SignOffRecord` are activities; experts, deployers, platform services, and institutions are agents. The value of this alignment is practical. Provenance is not merely a record of origin. It is a record that supports trust assessment, conflict reconstruction, and responsibility assignment. NIST's generative-AI profile likewise emphasizes content provenance, maintenance of change records with sources, timestamps, and metadata, TEVV record retention, and incident disclosure.

The operational rules are as important as the schema. Sampling should follow two channels. Random sampling is the baseline quality-control mechanism. It estimates silent error rates, deters complacency, and produces a representative stream of review data. Risk-weighted sampling is the safety mechanism. It prioritizes outputs whose domain, user context, uncertainty, or downstream action profile make harm more likely or more severe. NIST's GenAI profile recommends that risk management be tailored to context, severity, and deployment

conditions and that testing reflect real-world use rather than narrow benchmarks. EON therefore treats sampling as a dynamic control surface, not a fixed percentage.

Expert credentialing must be explicit and machine-checkable. At minimum, the network should verify identity, active license status where applicable, specialty, jurisdiction, disciplinary history if available, conflicts of interest, confidentiality commitments, and completion of EON-specific training on scope marking, evidence review, and audit obligations. Review eligibility should be bound to domain and jurisdiction; a clinician licensed in one jurisdiction should not silently sign off on practice that requires authorization in another. The EU AI Act's emphasis on competence, training, and authority for natural persons assigned to oversight points in the same direction, even though the Act does not define a professional review network.

Review workflows should be blind by default for independent review, structured by checklist, and time-bounded. The reviewer should see the relevant source-set, the claim graph, the system's uncertainty notation, and the current review state, but should not see other reviewers' initial judgments until the first-pass review is closed. This reduces conformity pressure and creates cleaner disagreement data. Review forms should support at least four actions: confirm, correct, reject, and mark outside scope. Where the system allows text edits, those edits must be stored as diffs attached to `UpdateEvent` objects rather than overwriting the initial record. NIST's AI RMF and SR 11-7 are aligned on this point: independent validation, critical review, documentation quality, and visible management responses are core determinants of sound governance.

Disagreement handling must preserve disagreement as data. If reviewers differ on a nonmaterial wording issue, the system may synthesize. If reviewers differ on a material claim, risk category, or recommendation, the system must open a `ConflictRecord`, mark the visible state accordingly, and escalate using preset rules. Escalation options include a senior reviewer, a second panel, or domain-specific blocking rules. The key requirement is that conflict not vanish in a consensus veneer. For professional systems, conflict is often epistemically meaningful. It tells the deployer, user, and insurer that the answer remained contested at release time.

Human sign-off thresholds should be narrow, explicit, and domain-bound. Not every professional output requires a named human signatory. But some do. The threshold should be crossed when the output either authorizes or materially conditions an action with direct effects on health, legal rights, financial exposure, or public rights and safety. In those cases, the sign-off record is not a ceremonial step. It is the state transition that identifies the final accountable decision node. That matches both the AI Act's stronger treatment of human oversight in high-risk settings and OMB's instruction that additional human oversight and intervention be required where significant impacts on rights or safety are possible.

The table below provides a practical default matrix for sampling and insurance posture.

Risk tier	Typical output class	Review mode	Default review rate	Human sign-off threshold	Insurer-relevant evidence
Low	General information, educational explanation, low-stakes summarization	Random post-delivery sampling	1–5%	None	Baseline error rate, audit completeness, incident reporting latency
Medium	Professional guidance without direct execution authority	Mixed random and risk-weighted review; provisional release allowed	5–15%	Required only when predefined triggers fire	Verified routing logic, reviewer eligibility, correction rate by category
High	Individualized advice with material risk if wrong	Pre-delivery review by one qualified expert; second reviewer on trigger	50–100% depending on domain policy	Usually required for release	Claim-level provenance, reviewer scope marks, override records, retention policy

Critical	Court filing, patient-specific treatment step, financial execution, public decision affecting rights or safety	Mandatory pre-delivery review and explicit sign-off; escalation if conflict	100%	Always required before release or execution	Full audit chain, named accountable human, immutable logs, documented fail-safe and incident response

These ranges are design defaults, not universal legal rules. Their value is that they create an underwriting surface. Underwriters do not need perfect safety. They need stable covariates: sampling rate, correction rate, credential coverage, conflict frequency, override frequency, time-to-correction, and record retention quality. Where those variables are absent, pricing becomes conservative or unavailable. Where they are present, insurability improves. That conclusion is an inference from the broader regulatory emphasis on logging, retention, independent evaluation, and model validation rather than an assertion that any regulator has already endorsed a specific EON schedule. The inference is nevertheless strong.

Incident Reconstruction and Audit Bloat

EON should not be misread as a demand for unlimited logging. More records do not automatically produce more accountability. Unstructured accumulation can create audit bloat: a condition in which prompts, model outputs, retrieved passages, expert comments, edits, metadata, and system events pile up until investigation becomes slower rather than clearer.

The goal is not maximal logging. The goal is reconstructable accountability.

A serious EON deployment should therefore distinguish between ordinary retention and incident reconstruction. For routine low-risk outputs, lightweight metadata may be sufficient: model version, time, domain label, risk tier, review state, and release state. For medium- and high-risk outputs, the system should preserve source records, claim records, review tasks, verification records, conflict records, update events, sign-off records, and audit-log entries. For critical outputs or disputed incidents, the system should generate an incident bundle: a frozen package

containing the relevant source objects, active memory state, claim graph, retrieval path, risk classification, expert reviews, conflict status, release state, user-visible output, human sign-off records, warnings, overrides, and downstream action history.

This distinction matters. In an accident, investigators do not need an undifferentiated ocean of logs. They need the event chain. A properly structured incident bundle should allow the system to reconstruct the path from source to claim, from claim to memory state, from memory state to AI output, from output to expert review, from review to release, and from release to user action. If that chain cannot be reconstructed, the system has not achieved traceable responsibility, no matter how much raw data it stores.

Retention should therefore be risk-tiered. Low-risk general information may require only lightweight operational records and aggregate sampling logs. Professional guidance should retain claim-level provenance and review status. High-risk individualized advice should retain reviewer scope, corrections, unresolved conflicts, and release authority. Critical outputs affecting health, legal rights, financial execution, public benefits, or physical safety should trigger full evidence preservation and incident-freeze capability.

This design also protects privacy and cost control. EON does not require every interaction to become a permanent dossier. It requires each interaction to carry enough structure to determine what must be preserved if risk increases. Record density should rise with consequence. The system should log enough to reconstruct the event and structure enough to avoid drowning in the log.

The difference is decisive. Without structure, more data creates more confusion. With structure, more data creates better accountability. EON therefore treats logging as a memory-governance function, not a data-hoarding reflex.

Liability, Insurance, and Domain Deployment

EON matters because it changes loss allocation from residual blame to structured responsibility. The provider remains responsible for the memory architecture, provenance integrity, routing logic, review-state labeling, credential verification controls, and tamper-evident logging. The deployer remains responsible for the chosen use case, threshold configuration, user communications, integration into the workflow, and any decision to bypass or disable a required review gate. The expert reviewer becomes responsible only within the scope actually reviewed and signed off. The user remains responsible for concealed facts, unauthorized use, or disregard of warnings. That layered approach is consistent with the move in recent governance documents away from vague responsibility language and toward explicit lifecycle roles, documentation, and control structures.

Insurability depends on evidence discipline. At minimum, an EON-enabled system intended for high-stakes deployment should have immutable or tamper-evident audit logs; version-pinned sources and model identifiers; claim-level provenance; reviewer credential records; contractually defined review states; signed incident-response procedures; retention schedules; and explicit authority rules for override and release. NIST's log-management guidance remains directly relevant here: an effective log infrastructure must support generation, transmission, storage, analysis, and disposal of log data; preserve original logs; support integrity checking; and meet retention and preservation requirements driven by legal or regulatory need. Those requirements map almost one-to-one onto the evidence needed by claims handlers, regulators, and counterparties after an AI incident.

Contractual controls should be equally explicit. Professional AI agreements should define reviewed and unreviewed states; the limits of review scope; mandatory review tiers; reviewer service-level expectations; confidentiality and conflict rules; error-correction and incident-notification procedures; audit rights for enterprise customers and insurers; retention periods; permitted downstream uses; and liability allocation for bypass, misuse, or unauthorized execution. Where licensed expert sign-off is part of the service, the contract should identify whether that sign-off is advisory, required for release, or required for execution. It should also specify which insurance tower responds first: technology E&O, cyber, professional liability, or a platform-issued combined policy. Those allocations are design proposals rather than

statements of current universal law, but they are the minimum structure needed for insurable professional deployment.

Domain deployment does not require different architecture, but it does require different thresholds.

In medicine, the minimum compliance posture should include clinician credential verification by specialty and jurisdiction; patient-data privacy controls; a clear distinction between informational content and patient-specific clinical recommendations; mandatory pre-delivery review for treatment recommendations, dose suggestions, contraindication-sensitive outputs, or triage decisions above a defined severity threshold; sign-off rules for outputs that would materially guide patient care; and incident escalation for adverse-event patterns. FDA's current CDS guidance remains important because it clarifies that some decision-support software functions fall outside device regulation while others remain regulated device functions; in either case, the provider must know which class of function it is operating and should not treat that classification question as an afterthought.

In law, the minimum compliance posture should include jurisdiction mapping; licensed-attorney review for court filings, legal opinions individualized to a client's facts, settlement or pleading drafts, and any output that asserts legal authority as a basis for action; citation verification before release; privileged-data controls; and retention of prompts, authorities, edits, and final release state. The ABA's first formal generative-AI guidance, as reported by Reuters, emphasizes competence, confidentiality, communication, fees, and the duty to review AI outputs for accuracy; it also notes that lawyers remain exposed to sanctions and misrepresentation risk if inaccurate AI-generated authorities are filed without verification. EON gives those obligations an operational form.

In finance, the minimum compliance posture should include independent validation of material models or workflows; separation between development and effective challenge; calibrated human approval thresholds for recommendations that exceed monetary or suitability limits; ongoing monitoring for drift; and documented overrides. Federal Reserve and OCC supervisory guidance on model risk management is still the best concise statement of the principle: validation should cover inputs, processing, and reporting; it should be independent from

development and use; it should continue after deployment; and significant deficiencies should block or tightly constrain use until resolved. Those principles fit naturally into EON review gates for advisory and decision-support systems.

In public administration, the minimum compliance posture should include an AI impact assessment, real-world-context testing, independent evaluation by a reviewing authority not directly involved in development, ongoing monitoring, additional human oversight and fail-safes for decisions affecting rights or safety, and centralized logging of waivers or exceptions. OMB Memorandum M-24-10 already requires those elements for covered federal uses of AI. EON is therefore best understood in the public sector as an implementation pattern for reviewable human judgment, not as a new normative claim.

Financing, Procurement, and Bankability

EON has financial consequences beyond insurance. A professional AI system with traceable expert oversight becomes easier to finance, purchase, and deploy inside institutions that must survive due diligence.

Banks and institutional lenders do not evaluate AI systems by technical capability alone. They evaluate loss predictability, contractual stability, operational controls, evidence preservation, and the chance that a borrower will face unbounded liability. A professional AI provider that offers high-impact outputs without source records, review trails, sign-off rules, incident bundles, or clear responsibility allocation remains difficult to finance. The lender cannot see where the risk begins, how it moves through the system, which actor absorbs which duty, or where the exposure stops.

EON changes that profile by turning expert oversight into financial evidence.

A lender can examine review coverage, risk-tier rules, expert qualification standards, correction rates, unresolved conflict rates, escalation protocols, audit-log completeness, insurance coverage, and incident reconstruction procedures. These are operational controls. They show whether the system has a working method for detecting error, routing judgment, preserving evidence, and limiting loss.

This matters for AI companies seeking debt financing or infrastructure-scale capital. Without a traceable responsibility structure, the company remains heavily dependent on equity financing, forward-looking valuation narratives, or strategic investors willing to absorb uncertainty. With EON, the company can present a more bankable operating model: professional outputs are risk-classified, high-risk cases are reviewed, critical cases require sign-off, expert corrections are logged, disputes are preserved, and failures can be reconstructed.

The same logic applies to enterprise procurement. Hospitals, law firms, banks, insurers, engineering firms, and public agencies do not buy high-risk AI merely because benchmark performance is strong. Their legal, compliance, finance, risk, and insurance teams must approve deployment. EON gives those teams something concrete to inspect. They can ask: What percentage of outputs are reviewed? Which outputs require mandatory review? What qualifications do reviewers hold? How are disagreements handled? What records survive after an incident? Which liabilities remain with the vendor, deployer, expert reviewer, and end user?

A generic human-in-the-loop promise cannot answer those questions. EON can, because it binds human judgment to records, credentials, review states, escalation rules, and sign-off authority.

This is why expert oversight should be understood as commercial infrastructure. It makes AI professional services easier to contract, audit, finance, insure, and scale. A buyer purchases more than model access; it purchases an operating structure with defined controls, review states, evidence trails, and responsibility boundaries.

EON therefore changes the financial status of professional AI. It moves AI systems from venture-funded capability demonstrations toward institutional-grade service infrastructure. Capability attracts attention. Traceable judgment attracts capital that must survive due diligence.

Once expert judgment becomes traceable, professional AI becomes more insurable, more purchasable, and more bankable.

Implementation, Conclusion, and Policy Recommendations

A practical implementation path begins with narrow scope. The first milestone is not full expert-marketplace scale. It is the creation of the memory objects defined above, because without claim-level provenance and update events there is nothing reliable to review. The second milestone is a credentialing and routing service that can bind reviewers to domain, jurisdiction, and conflict rules. The third milestone is a pilot in one domain with two review modes only: random sampling for low-risk outputs and mandatory pre-delivery review for a strictly defined high-risk subset. The fourth milestone is a visible review-state label and audit portal for enterprise customers. The fifth milestone is an incident-response and insurer evidence package. The sixth milestone is expansion from one domain to multi-domain operations only after correction rates, reviewer disagreement rates, and latency targets stabilize. This sequence follows the same logic that NIST, OMB, and SR 11-7 all favor: governance, documentation, testing, independent review, and continuous monitoring before scale.

The practical priority order is clear. Build the records first. Then build the review gates. Then build the contractual and insurance interfaces. Many AI providers attempt the reverse order. They market reliability, promise safety, and negotiate liability carve-outs before they can reconstruct how a consequential output reached its final form. That sequence is unstable. EON inverts it. It treats professional review as a state machine with evidence, not as an assurance slogan.

The conclusion is limited but strong. Vol. 1 supplied the technical substrate for Presence. Vol. 2 supplied the institutional substrate for Integrity. Vol. 3 supplies the professional verification layer that makes those substrates usable in high-stakes domains. Expert judgment only strengthens AI governance when it is bound to provenance, written into memory, exposed through review states, and tied to a responsibility architecture. Otherwise, “human oversight” remains mostly ceremonial. EON is valuable because it makes human judgment locatable.

The policy recommendations follow directly. Regulators should recognize traceable, risk-calibrated expert review as a valid form of effective human oversight where the system preserves logs, identifies review scope, and defines override authority. Providers of high-risk professional AI should expose machine-readable review states to downstream actors rather than presenting all outputs as equivalent. Procurement rules and enterprise contracts should

require access to audit records sufficient to reconstruct source, review, conflict, update, and sign-off history. Professional bodies should define minimum competency, conflict, and jurisdiction rules for AI oversight work, so that “expert reviewer” becomes a governed role rather than an informal label. Insurers should condition coverage improvements on evidence quality rather than on marketing claims about alignment or model character. Those recommendations are modest. They do not require fully interpretable models. They require a system in which responsibility can move through records rather than through guesswork.

References

Guo, Xiangyu. *Frame Theory for Artificial Intelligence: Presence × Integrity as the Structural Law of Trust and Governance*. 2025. Zenodo

Guo, Xiangyu. *Frame Infrastructure for AI, Vol. 1: Beyond Human-Like Memory — Toward a Machine-Native Architecture of Structural Persistence*. Zenodo, 2025. DOI: 10.5281/zenodo.20738963.

Guo, Xiangyu. *Frame Infrastructure for AI, Vol. 2: From Black-Box Safety to Traceable Responsibility — Machine-Native Memory, Liability Allocation, and the Insurability of AI Risk*. Zenodo, 2025. DOI: 10.5281/zenodo.20738983.

Guo, Xiangyu. *Expert Oversight Network (EON): A Paradigm for Trustworthy AI-Driven Professional Advice*. Pre-existing article, 2025.

NIST. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1, 2023.

NIST. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1, 2024.

World Wide Web Consortium. *PROV-DM: The PROV Data Model* and *PROV-Overview*. 2013.

European Union. Regulation (EU) 2024/1689, *Artificial Intelligence Act*. Official Journal of the European Union, 2024.

NIST. *Guide to Computer Security Log Management*. SP 800-92, 2006.

Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency. *Supervisory Guidance on Model Risk Management*. SR 11-7, 2011.

U.S. Food and Drug Administration. *Clinical Decision Support Software: Guidance for Industry and Food and Drug Administration Staff*. 2026.

Office of Management and Budget. Memorandum M-24-10, *Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence*. 2024.

Merken, Sara. “Lawyers using AI must heed ethics rules, ABA says in first formal guidance.” *Reuters*, 2024.