WHITE PAPER ·
THE UNIVERSITY OF TEXAS AT DALLAS X ARTSCILAB

Council of LLMs:

A Multi-Agent AI Architecture for Legal Reasoning

Adversarial Deliberation · Structured Epistemic Output · Hallucination-Resistant Design

LEAD AUTHOR:

Venkatesh Prasad Ravichandran

Master of Science, Business Analytics & Artificial Intelligence

The University of Texas at Dallas

Co Author:

Kiran Akshay Sundhararaajan

Masters of Science, Computer Engineering

University of Texas at Dallas

And yes, I had help from the best AI models out there

REVIEWED & CREDITED BY

Conceptual Foundation & Original Vision

The Honorable Don John McClellan Marshall

Senior United States District Judge (Ret.) ArtSciLab, The University of Texas at Dallas

REVIEWED & CREDITED BY

System Design & Engineering Review

Dr. Paul Fishwick

Distinguished University Professor of Arts and Technology Dept. of Arts and Technology · Erik Jonsson School The University of Texas at Dallas

April 2026 · ArtSciLab · The University of Texas at Dallas

Abstract

This white paper presents Council of LLMs, a novel multi-agent artificial intelligence architecture for rigorous, structured legal reasoning. Unlike single-model legal AI tools, Council of LLMs assembles a deliberative body of specialized AI agents each constrained to a distinct institutional role that argue, cross-examine, and rule on legal questions in a manner that mirrors the adversarial structure of common law.

The system is grounded in a seven-phase pipeline that moves from query intelligence through parallel research, adversarial advocacy, cross-examination, a specialist judge panel, a streaming chief judge ruling, and post-ruling verification. The architecture is built around a single non-negotiable priority: the elimination of hallucinated legal authority. Every cited case, statute, and proposition is traced to a verified source in the research bundle before it may appear in any ruling.

SECTION 1

Introduction: The Problem with Single-Voice Legal AI

The Structural Flaw

Legal reasoning is not monolithic. A legal question submitted to a single AI model receives a single answer shaped by a single perspective, a single training distribution, and a single set of learned biases. This is precisely the opposite of how the law operates.

The common law tradition has, for centuries, resolved uncertainty through structured adversarial argument: a plaintiff's counsel who argues as hard as possible for one position, a defense counsel who argues as hard as possible for the opposing position, and a neutral judge who evaluates both arguments against the law and the facts. This structure exists not as procedural formality, but as an epistemological method a systematic procedure for surfacing the strongest possible arguments on all sides before a decision is made.

When a single AI model is asked a legal question, it performs an informal internal simulation of this process one that is invisible, unverifiable, and unaccountable. The model's internal deliberation is opaque. Its cited cases may or may not exist. There is no adversarial pressure that forces the strongest counterargument to be stated.

> *"A system that presents conclusions with apparent authority but without the structural integrity that gives legal authority its meaning is not a legal tool it is a liability."*

The Council of LLMs Response

Council of LLMs was designed to solve this problem by externalizing and structuring the deliberative process that legal reasoning actually requires. Rather than asking one model to reason about a legal question from one vantage point, the system assembles a council a structured set of agents with distinct, institutionally constrained roles that conduct the full adversarial process in a form that is visible, traceable, and auditable.

The result is not just an answer. It is a ruling: a structured document that states what the facts are and which are disputed, which legal standards apply, what the strongest arguments on each side are, which precedents are most relevant, and what conclusion flows from applying the law to the facts with every proposition grounded in a verifiable source.

SECTION 2

Vision: The Deliberative Council Model

The vision behind Council of LLMs draws directly from the architecture of judicial deliberation. A well-functioning court is not a single voice. It is a structured institution with defined roles advocates who argue, witnesses who testify, judges who decide and defined rules about what kinds of evidence and argument count.

Council of LLMs translates this institutional architecture into an AI system:

- Advocates argue for each side, building the strongest possible case from available evidence and law.
- A cross-examination phase forces each advocate to confront the weaknesses in its own position and the strengths of the opposing position.
- Specialist sub-judges independently evaluate disputed facts, applicable legal standards, and analogous precedents.
- A Chief Judge synthesizes everything into a formal ruling with explicit epistemic markers.
- Post-ruling verification checks every cited authority against the verified research bundle before the ruling is finalized.

Why This Matters

Legal questions arise constantly in contexts where access to sophisticated legal analysis is limited. Small businesses, individuals navigating the civil system, public interest organizations, and legal aid clinics face legal questions whose complexity exceeds the resources available to answer them properly. A system that can conduct rigorous, multi-perspective legal analysis and be transparent about the limits of its conclusions democratizes access to the kind of analysis that sophisticated parties take for granted.

SECTION 3

Philosophical Foundations

Adversarialism as Epistemology

The adversarial structure of the common law is not merely a procedural convention. It reflects a deep epistemological commitment: that the best way to discover the truth about a contested question is to force the strongest possible argument on each side to compete openly, with an impartial decision-maker evaluating the result.

Karl Popper's philosophy of falsification offers a useful analogy. Scientific knowledge advances not by accumulating confirmatory evidence but by subjecting theories to the strongest possible attempts at refutation. Theories that survive serious attempts to disprove them are more reliable than theories that have never been challenged. The adversarial system applies the same logic to legal fact-finding and legal reasoning.

Epistemic Humility as a Design Constraint

The second philosophical commitment is epistemic humility the recognition that not all legal questions have clear answers, and that a system that presents uncertain conclusions with false confidence is more dangerous than one that acknowledges uncertainty. The system implements epistemic humility through structural mechanisms rather than relying on model self-assessment.

Every claim in the final ruling is tagged with an epistemic status derived from the structured pipeline data: SETTLED (all agents agree, multiple sources), DISPUTED (agents disagreed), or SINGLE SOURCE (supported by only one source). A consensus level from STRONG CONSENSUS to FUNDAMENTAL DISAGREEMENT is computed from structural pipeline data not from any model's self-reported confidence.

The Role of Multiple Minds

The use of multiple distinct AI models rather than a single model, or multiple instances of the same model reflects a third philosophical commitment: that epistemic diversity produces more reliable conclusions than epistemic monoculture. Different large language models have different training distributions, different architectural biases, and different reasoning tendencies.

SECTION 4

The Adversarial Method Applied to AI

Why Advocates, Not Assistants

Most AI legal tools are designed to be helpful assistants: they answer questions, summarize documents, and explain the law. These tools share a structural limitation they produce analysis from a single perspective, optimized for helpfulness rather than for surfacing the best counterargument. Council of LLMs is designed around advocates, not assistants.

The Build → Critique → Refine Chain

Each advocate agent operates through a three-step chain:

- Build The advocate constructs its initial argument, citing sources from the research bundle and applying the applicable legal standards to the facts as it understands them.
- Critique A separate critic agent (Haiku 4.5) reads the argument and identifies factual inaccuracies, unsupported claims, and logical weaknesses as a grounding check against the research bundle.
- Refine The advocate revises its argument in light of the critique, strengthening weak points and correcting any factual inaccuracies.

Cross-Examination

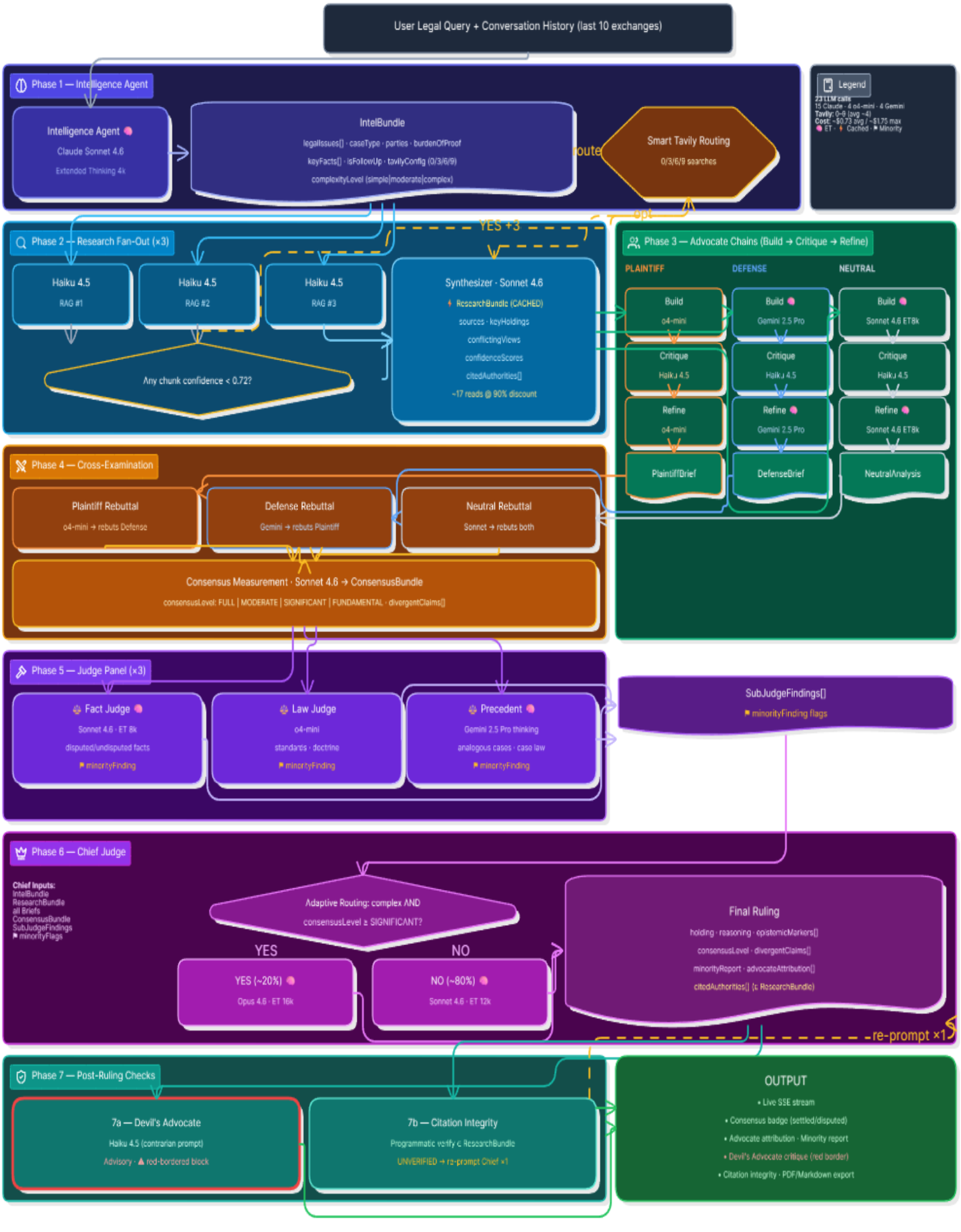
After the advocate chains complete, each advocate is given the opposing arguments and asked to rebut them. The output includes not only the rebuttals but also a structured consensus measurement ranging from STRONG CONSENSUS to FUNDAMENTAL DISAGREEMENT which determines which model handles the final ruling and how much extended reasoning budget is allocated to it.

SECTION 5

System Architecture

Council of LLMs is implemented as a seven-phase sequential pipeline. The output of each phase is structured JSON that becomes typed input to the next phase. No phase may invent information that was not produced by a prior phase.

Council of LLMs — Legal AI Pipeline



The pipeline is served over Server-Sent Events (SSE), allowing the client to display real-time phase progress, streaming token output from the Chief Judge, and the final verified ruling all in a single long-lived connection.

Data Flow Between Phases

FROM PHASE	OUTPUT BUNDLE	TO PHASE
1 Intelligence Agent	IntelBundle (caseType, issues, complexity, tavilyConfig)	2
2 Research Fan-out	ResearchBundle (sources with stable src-N IDs)	3, 4, 5, 6
3 Advocate Chains	3× AdvocateResult (argument, keyPoints, citations)	4
4 Cross-Examination	CrossExam (rebuttals, consensusLevel, divergentClaims)	5, 6
5 Judge Panel	SubJudgeFindings (fact, law, precedent) + minorityFlag	6
6 Chief Judge	FinalRuling (full schema, streamed)	7
7 Verification	VerifiedRuling + DA Critique + CitationReport	Client

SECTION 6

Phase-by-Phase Methodology

Phase 1 Intelligence Agent

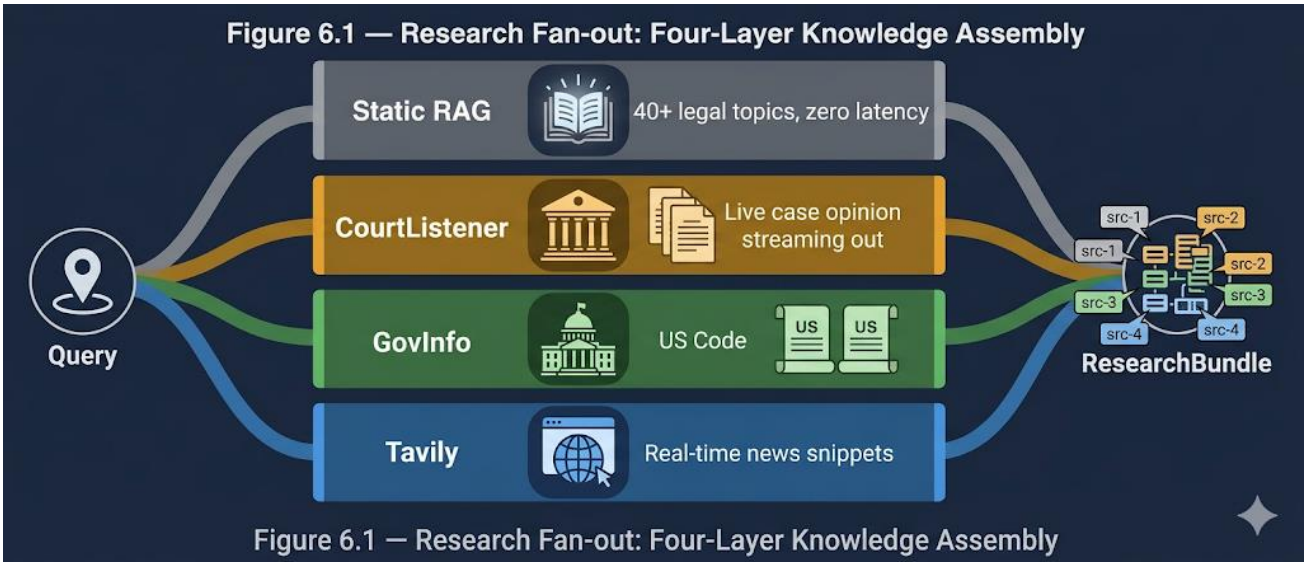
Model: Claude Sonnet 4.6 with extended thinking (adaptive budget: 3,000–12,000 tokens).

The Intelligence Agent is the entry point of the pipeline. It reads the raw user query and conversation history and produces a structured IntelBundle: legalIssues[], caseType, parties, burdenOfProof, keyFacts[], jurisdiction, complexityLevel, and tavilyConfig. The tavilyConfig implements smart search routing stable areas of law receive zero web searches; volatile areas receive up to nine.

Phase 2 Research Fan-out

Models: Three Claude Haiku 4.5 instances (parallel research) + Claude Sonnet 4.6 (synthesis).

Four source layers are fetched simultaneously. Three Haiku instances conduct parallel research from each advocate perspective. Their outputs are merged by a Sonnet 4.6 synthesizer that produces the final ResearchBundle a structured document with all sources assigned stable src-N IDs. These IDs are the foundation of citation integrity throughout the pipeline.



Phase 3 Advocate Chains

Models: OpenAI o4-mini (Plaintiff) · Gemini 2.5 Pro (Defense) · Claude Sonnet 4.6 (Neutral).

Three advocate chains run in parallel, each executing the Build → Critique → Refine sequence. The assignment of distinct reasoning model architectures to the three advocate roles creates genuine analytical diversity the three advocates are not running the same computation with different instructions; they are running genuinely different computations. Each advocate's output contains argument, keyPoints[], and citations[] referencing only ResearchBundle source IDs.

Phase 4 Cross-Examination + Consensus Measurement

Models: Claude Haiku 4.5 ×3 (rebuttals) + Claude Sonnet 4.6 (consensus measurement).

Each advocate is given the opposing arguments and asked to identify their weaknesses. Sonnet 4.6 then produces a structured consensus assessment consensusLevel, divergentClaims[], and agreedFacts[]. The consensusLevel has operational consequences: it determines the model and thinking budget for Phase 6.

Phase 5 Specialist Judge Panel

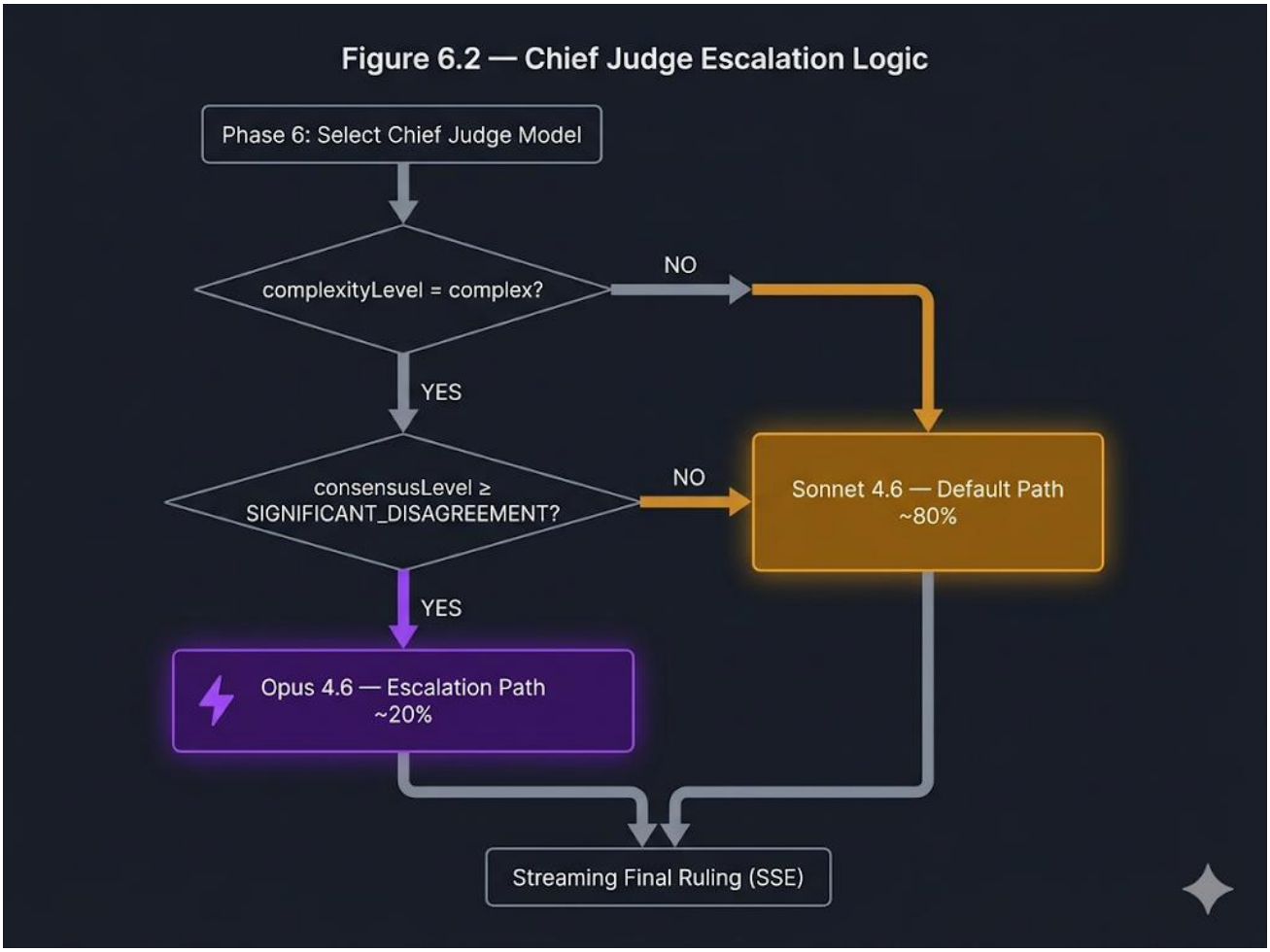
Models: Claude Sonnet 4.6 (Fact Judge) · OpenAI o4-mini (Law Judge) · Gemini 2.5 Pro (Precedent Judge).

Three specialist evaluations run in parallel, each strictly limited to its designated domain. The Fact Judge determines only what is factually contested. The Law Judge identifies only the doctrinal framework. The Precedent Judge identifies analogous cases and their overall weight. A minorityFinding flag is set when sub-judges diverge significantly.

Phase 6 Chief Judge (Streaming)

Models: Claude Sonnet 4.6 + thinking (default, ~80% of queries) · Claude Opus 4.6 + max thinking (escalation, ~20%).

The Chief Judge receives the complete context from all prior phases and produces the final ruling as structured JSON, streamed token-by-token to the client via SSE. Escalation to Opus 4.6 occurs when both conditions are met: complexityLevel = complex AND consensusLevel is SIGNIFICANT_DISAGREEMENT or FUNDAMENTAL_DISAGREEMENT. Every citedAuthority must reference a src-N ID from the ResearchBundle.



Phase 7 Post-Ruling Verification

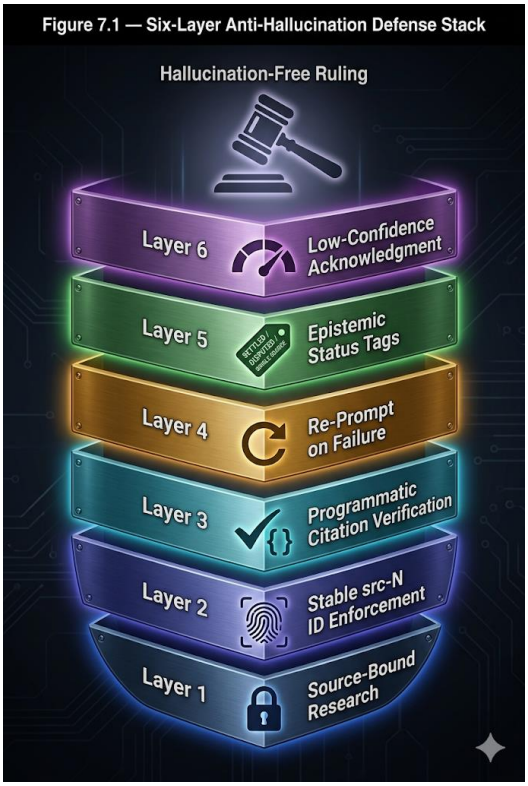
Models: Claude Haiku 4.5 (Devil's Advocate) + deterministic programmatic check (Citation Integrity).

7a: Devil's Advocate Haiku 4.5, under an explicitly contrarian system prompt, identifies assumptions, missed arguments, systematic biases, and overconfident claims. Displayed in a red-bordered advisory block. 7b: Citation Integrity A deterministic program verifies every cited src-N ID against the ResearchBundle. On failure, the system re-prompts once with the corrected ID list. No hallucinated citation silently survives into the final output.

SECTION 7

Anti-Hallucination Architecture

Hallucinated legal citations are among the most dangerous failure modes in AI legal tools. Council of LLMs addresses this at six distinct layers:

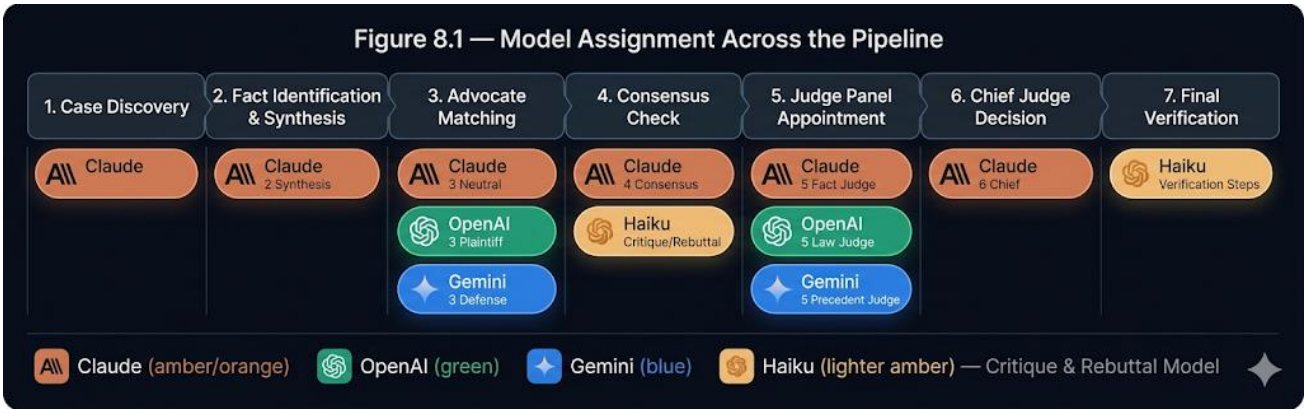


LAYER	MECHANISM	EFFECT
1. Source-bound research	ResearchBundle assembled from verified sources before any advocate runs	Agents cannot cite non-existent sources
2. Stable ID enforcement	All sources assigned src-N IDs; downstream agents given the valid list	IDs are traceable and finite
3. Programmatic verification	Phase 7b checks every citation mechanically, no model behavior relied upon	Hallucinated IDs cannot survive silently
4. Re-prompt on failure	One re-prompt opportunity with corrected ID list before failure is reported	One chance for model self-correction
5. Epistemic markers	Claims tagged settled / disputed / single_source from pipeline structure	Evidentiary thinness made visible
6. Low-confidence acknowledgment	Chief Judge instructed to state low confidence when evidence is thin	Structural humility, not prompt-dependent

SECTION 8

Model Selection and Epistemic Diversity

The choice of which models play which roles reflects two considerations: matching model capabilities to task requirements, and achieving genuine diversity of reasoning perspective across the advocate and judge roles.



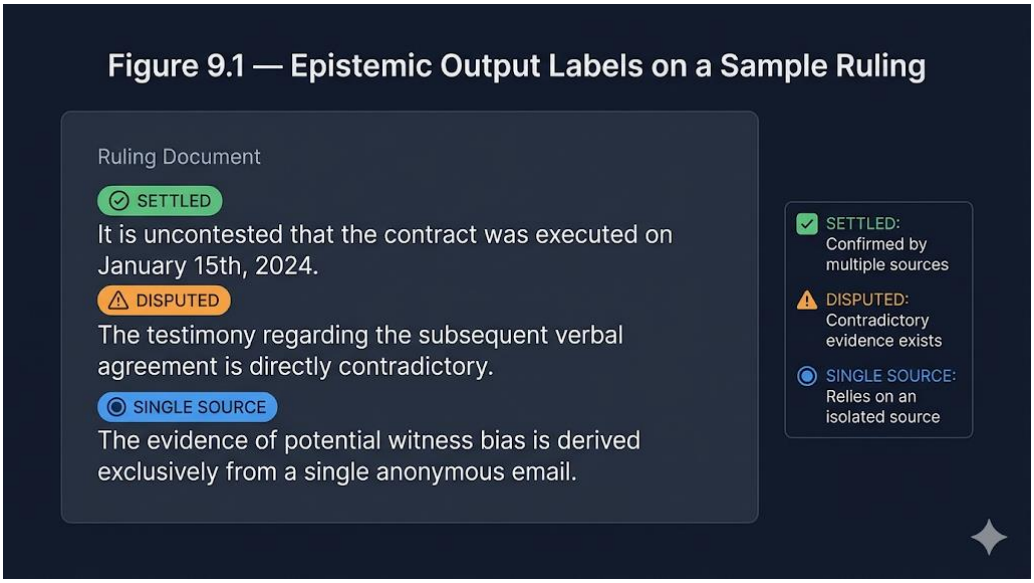
PHASE	AGENT	MODEL	RATIONALE
1	Intelligence Agent	Claude Sonnet 4.6 + thinking	Extended thinking for nuanced query classification
2	Advocate researchers	Claude Haiku 4.5 ×3	High throughput, low cost, parallel
2	Synthesizer	Claude Sonnet 4.6	Quality synthesis of multi-source research
3	Plaintiff advocate	OpenAI o4-mini	Chain-of-thought reasoning; different training lineage
3	Defense advocate	Gemini 2.5 Pro	Thinking model; architectural diversity
3	Neutral advocate	Claude Sonnet 4.6 + thinking	Extended thinking for procedural precision
3	All critics	Claude Haiku 4.5	Fast factual grounding check
4	All rebuttals	Claude Haiku 4.5	Speed and cost efficiency
4	Consensus measurement	Claude Sonnet 4.6	Quality classification of divergence
5	Fact Judge	Claude Sonnet 4.6 + thinking	Extended thinking for fact-law separation
5	Law Judge	OpenAI o4-mini	Reasoning model for doctrinal precision
5	Precedent Judge	Gemini 2.5 Pro	Large context for case pattern matching
6	Chief Judge (default)	Claude Sonnet 4.6 + thinking	High quality, cost-effective (~80%)
6	Chief Judge (escalation)	Claude Opus 4.6 + max thinking	Maximum capability for hardest cases (~20%)

PHASE	AGENT	MODEL	RATIONALE
7	Devil's Advocate	Claude Haiku 4.5	Fast contrarian critique
7	Citation check	Deterministic (no model)	Zero model hallucination risk

SECTION 9

Structured Output and Epistemic Transparency

Every inter-phase boundary in the pipeline is a typed JSON schema. This is not a convenience it is a correctness constraint. Epistemic signals are structural artifacts of the pipeline, not model-generated self-reports.



A model asked "how confident are you?" will produce a response calibrated to what sounds appropriate. A system designed to track disagreement structurally and surface it explicitly does not need to ask. Three structural outputs guarantee this:

- No information invention at boundaries the schema defines the universe of information downstream agents may rely on.
- Verifiable outputs every structured output can be programmatically inspected; citation checking is possible because citations are in a typed array, not buried in prose.
- Derived epistemic signals consensusLevel, epistemicMarkers, minorityReport, and advocateAttribution are all derived from structured pipeline data produced by prior phases.

SECTION 10

Knowledge Infrastructure

Static Knowledge Base (RAG)

The system maintains an in-memory legal knowledge base covering over forty discrete topics across eleven areas of law: property, contract, tort, employment, constitutional, criminal procedure, civil procedure, family, consumer protection, intellectual property, and general procedure and evidence. Each entry includes detailed content with specific statutory citations and landmark case references, and keywords for scored relevance matching.

Live Legal Databases

SOURCE	COVERAGE	ROUTING STRATEGY
CourtListener	Federal and state appellate opinions, continuously updated	Routed to courts most relevant per case type (e.g., CAFC for IP, CADC for regulatory)
GovInfo	United States Code statutory text	Direct package lookup by pre-mapped USC title per case type no search API, reliable GET
Tavily	Real-time web recent rulings, regulatory updates, new legislation	0–9 searches; conditional on domain volatility score from Intelligence Agent

SECTION 11

Performance and Cost Design

COMPONENT	CALLS / QUERY	NOTES
Claude (Haiku + Sonnet + Opus)	Up to 15	All phases
OpenAI o4-mini	4	Plaintiff advocate + Law Judge
Gemini 2.5 Pro	4	Defense advocate + Precedent Judge
TOTAL	~23	Per full pipeline execution

Cost Optimization Strategies

- Prompt caching ResearchBundle wrapped with cache_control: ephemeral on Anthropic models, enabling ~90% cost reduction across ~17 downstream reads. The single largest optimization in the system.
- Adaptive thinking budgets complexityLevel from Phase 1 drives budget allocation. Simple: 3,000 tokens. Complex: 12,000–16,000 tokens. Maximum reasoning is never applied to questions that don't require it.
- Model tiering Haiku 4.5 handles all critique, rebuttal, and verification. Sonnet handles synthesis and evaluation. Opus is reserved for escalation-path (~20% of queries).
- Selective web search Zero-to-nine Tavily routing eliminates latency and cost on stable-law queries.

COST PATH	ESTIMATED COST	CONDITIONS
Minimum	~$0.30	Simple query, zero Tavily searches, Sonnet Chief Judge
Average	~$0.73	Moderate query, 4–5 Tavily searches, Sonnet Chief Judge
Maximum	~$1.75	Complex query, 9 Tavily searches, Opus escalation path

SECTION 12

Use Cases and Scope

AUDIENCE	USE CASE
Pro se litigants	Understand legal framework, applicable standards, and strongest arguments on both sides before deciding how to proceed
Legal aid organizations	Rapidly assess legal merits of potential matters at intake before committing attorney time
Law students & researchers	Trace how adversarial argument develops across a legal question with explicit sourcing
Experienced practitioners	Rapid research across unfamiliar practice areas; stress-test draft arguments before filing
Policy researchers & journalists	Understand the legal landscape around regulatory questions with explicit sourcing

SCOPE LIMITATION: Council of LLMs is not a substitute for a licensed attorney. It does not create an attorney-client relationship.

Its analysis is a structured starting point for legal inquiry, not a professional legal opinion.

Users must verify all cited authorities and consult licensed counsel before taking legal action.

SECTION 13

Limitations and Ethical Considerations

Jurisdictional Coverage

The system's static knowledge base and research routing are focused on United States federal law and selected state law, with particular depth in California. International and non-US legal questions are outside the current scope.

Training Data Cutoffs

All language models in the pipeline have training data cutoffs. The Tavily layer mitigates this for high-volatility domains, but the system may not reflect the most recent judicial decisions or statutory amendments in all areas. Users should independently verify the currency of cited authority.

Hallucination Risk

Despite the multi-layer anti-hallucination architecture, the system cannot guarantee that every cited proposition is correctly stated or that no fabricated authority survives verification. The citation integrity check catches hallucinated IDs but cannot catch mischaracterizations of correctly-cited sources. Legal research must always be independently verified.

Fairness and Bias

Language models reflect the statistical patterns of their training data, which includes historical legal texts and judicial opinions that embody the biases of the legal system itself. The system's multi-model, adversarial design reduces but does not eliminate the risk that any single model's biases dominate the analysis.

SECTION 14

Conclusion

Council of LLMs represents a principled attempt to align the architecture of AI legal analysis with the structure of legal reasoning itself. Where most AI legal tools offer a single perspective from a single model, this system offers a deliberative process: structured adversarial argument, cross-examination, specialist evaluation, and a final ruling that is explicit about what it knows and what it doesn't.

The design is grounded in three philosophical commitments: adversarialism as epistemology, epistemic humility as a design constraint, and epistemic diversity as a reliability mechanism. The anti-hallucination architecture source-bound research, stable ID enforcement, programmatic citation verification, and structured epistemic markers reflects the judgment that confident-sounding errors are more dangerous than acknowledged uncertainty.

> *"Rigorous, adversarially stress-tested legal analysis has historically been available only to parties with resources sufficient to hire teams of lawyers and researchers. A system that externalizes and structures this process in a transparent, auditable form extends that quality of analysis to any user who can formulate a legal question."*

[REPLACE WITH IMAGE]

Figure 14.1 Closing Visual: Justice Made Accessible

IMAGE PROMPT: *Wide-angle cinematic illustration: a lone figure stands at the base of a towering classical courthouse, looking up at the entrance; above the doors, instead of stone engravings, holographic glowing text reads "Council of LLMs"; the sky transitions from dark stormy gray on the left to warm golden sunrise on the right, symbolizing democratization of access to justice; seven glowing orbs (representing the seven AI agents) float in formation above the figure, casting light downward; epic scale, photorealistic concept art, dramatic warm-cool lighting contrast, ultra-detailed architecture, 8K*

Size / style note: Full width · 21:9 cinematic ratio · photorealistic epic concept art

SECTION 15

Future Work

DIRECTION	DESCRIPTION
Extended jurisdictional coverage	Expand static knowledge base and retrieval routing to cover additional state jurisdictions and international legal systems
Document-grounded analysis	Allow users to submit contracts, pleadings, and statutes as part of the query, incorporated into the ResearchBundle
Procedural guidance layer (Phase 8)	Translate the ruling into concrete procedural guidance forms to file, deadlines, next steps tailored to jurisdiction and role
Confidence calibration study	Systematic evaluation of the correlation between consensus levels and actual legal certainty as assessed by licensed attorneys
Human-in-the-loop integration	Mechanisms for licensed attorneys to review and annotate system outputs before presentation to end users
Longitudinal citation tracking	Integration with citation tracking services to flag when a cited precedent has been overruled, distinguished, or limited
Expanded model diversity	Systematic evaluation of new reasoning models as they become available to maximize genuine analytical diversity