Tribunal Session Summary

Session: conclave-20260514-002447-what-are-the-main-a9a9

Depth: T6

Advocates: 6 | Judges: 5 | Cost: $4.8377 | Time: 29m 21s

THE PROMPT: What are the main reasons people hire Harvey.ai? Include key benefits and any common use cases.

THE RULING: People hire Harvey.ai primarily to recover measurable attorney time—2 to 10 hours per week per lawyer—through two core use cases: first-draft generation of legal documents and fast contract/knowledge querying. Legal-domain specialization helps Harvey get shortlisted, but productivity gains close the deal. These benefits are enabled by Harvey's legal-specific design and ability to train on proprietary repositories, while mandatory lawyer oversight is a compliance constraint, not a driver.

The Prompt

What are the main reasons people hire Harvey.ai? Include key benefits and any common use cases.

The Ruling

People hire Harvey.ai primarily to recover measurable attorney time—2 to 10 hours per week per lawyer—through two core use cases: first-draft generation of legal documents and fast contract/knowledge querying. Legal-domain specialization helps Harvey get shortlisted, but productivity gains close the deal. These benefits are enabled by Harvey's legal-specific design and ability to train on proprietary repositories, while mandatory lawyer oversight is a compliance constraint, not a driver.

The strongest evidence is Harvey's own reporting that its internal legal team saves 20–40 hours per week and customers report 2–10 hours saved per attorney weekly, tied directly to drafting and querying workflows. The main caveat is that these figures are self-reported and unaudited, with no independent verification of whether they are net of review time or how Harvey compares to competitors. The tribunal converged on this answer because all advocates, after rigorous debate, centered their final positions on productivity as the primary driver, while rejecting unsupported claims about technical architecture, error reduction, and vendor culture.

Core Analysis

Thesis: Productivity Gains Are the Primary Hiring Driver

Organizations hire Harvey.ai not for speculative transformation but for measurable, repeatable recovery of attorney capacity. The core value proposition is **time savings**—specifically, the ability to reclaim 2 to 10 hours per attorney per week on high-volume, repetitive legal tasks. This is not a vague promise of "AI efficiency" but a targeted ROI tied to two dominant workflows: **first-draft generation** and **contract/knowledge querying**. These use cases are explicitly documented in Harvey's own materials and corroborated by third-party legal industry coverage.

The 2–10 hours/week figure is drawn from customer anecdotes reported in Harvey's 2024-08-19 blog, "How Harvey Uses Harvey," and is supported by the internal legal team's report of saving 20–40 hours per week. While these numbers are self-reported and unaudited, they are the only quantified productivity metrics in the record and are consistently cited across all judicial opinions. Even at a 50% discount for selection bias, the residual time savings represent material capacity gains—equivalent to adding 0.5 to one full-time attorney per team of five—making it the primary justification on procurement decks.

Key Use Cases: Drafting and Querying

First-Draft Generation of Legal Documents

Harvey is used to generate initial drafts of common legal documents such as Non-Disclosure Agreements (NDAs), contractor agreements, internal policies, and guidelines. These drafts are written in "easy-to-understand language that doesn't contain a lot of legalese," as stated in the 2024-08-19 Harvey blog. This reduces the time attorneys spend on boilerplate and repetitive drafting, compressing the cycle from blank page to reviewable draft.

The output is not intended to be final; it is a strong starting point that lawyers then edit and refine. This workflow is particularly valuable for in-house legal teams managing high volumes of routine agreements. The Maryland State Bar Association's 2024-02-22 overview confirms that drafting is a typical use case for legal AI tools, and Harvey's ability to be trained on a firm's proprietary document repositories enhances the relevance and consistency of these drafts.

Contract Querying and Knowledge Retrieval

Legal teams use Harvey to rapidly search and extract information from large volumes of executed contracts—over 1,000 in some cases—including Data Processing Agreements (DPAs), Business Associate Agreements (BAAs), order forms, and platform agreements. This allows lawyers to answer complex, deal-specific questions in seconds rather than hours, significantly accelerating due diligence, compliance checks, and contract lifecycle management.

For example, a lawyer can query Harvey to find all instances where a specific indemnity cap was negotiated or to surface the renewal terms across hundreds of vendor agreements. This use case is particularly valuable for corporate legal departments managing sprawling contract portfolios. The 2024-08-19 blog explicitly describes this as a "major use case" and a key contributor to the reported time savings.

Enabling Capabilities: Legal-Specific Design and Proprietary Training

While productivity gains close the deal, **legal-domain specialization** is what gets Harvey into the room. Unlike generic large language models (LLMs), Harvey is "specifically designed for legal professionals" (Maryland State Bar Association, 2024-02-22) and can be trained on a legal team's proprietary document repositories. This enables tailored workflows and ensures outputs are contextually and legally relevant.

One judicial opinion usefully frames this as a **"shortlisting vs. closing" mechanism**: legal specialization acts as a filter that signals fitness for purpose, making Harvey a credible option in a risk-averse industry. However, no judge found buyer-side evidence that this specialization is a *differentiating* factor among legal AI tools—only that it is a *prerequisite* for consideration. The ability to integrate with proprietary data enhances utility but is not the primary procurement driver.

Adoption Constraint: Lawyer Oversight Is a Requirement, Not a Benefit

Harvey explicitly states that its AI tools "are not meant to be used without lawyer review and oversight" and should not make unsupervised decisions (Harvey blog, 2024-08-19). This is not a selling point but a **compliance necessity** rooted in professional responsibility rules. All judges converged on the view that oversight is a *constraint* that enables adoption in regulated environments, not a value proposition that drives hiring decisions.

One advocate initially claimed oversight "shifts malpractice risk," but this was correctly rejected as unsupported and logically flawed—liability remains with the attorney regardless of AI use. The tribunal affirms that Harvey's positioning as a tool for augmentation, not replacement, is essential for trust and adoption, but it is not a primary reason for procurement.

Tertiary or Niche Factors: Mission and Vendor Approach

Access-to-Justice Mission

Harvey's long-term mission to "increase access to justice" (Deal Director / InfraPlay, 2023-11-30) and its partnership with Singaporean courts and educational institutions are real and cited in the record. However, these are niche drivers, relevant primarily in public-sector or pro bono contexts. No evidence suggests this mission influences corporate legal procurement decisions at scale.

Consultative Vendor Approach

Harvey seeks employees who are "truly consultative and intellectually curious" (Harvey blog, 2024-07-15), and some advocates suggested this customer-centric posture could be a differentiator. However, no buyer-side evidence supports this as a hiring driver. The tribunal treats this as a plausible but unverified secondary enabler of customer success, not a primary reason for adoption.

Rejected Claims

Several claims were explicitly rejected due to lack of evidence:

- **Document management capabilities**: Not mentioned in any source; dismissed as fabricated.
- **Error reduction or negotiation speed**: No evidence in the record; advocates who made these claims conceded them under challenge.
- **Plain-language outputs accelerate editing**: While Harvey produces "easy-to-understand language," no source measures editing time. In fact, one judge noted that removing legalese may require lawyers to re-insert precise legal terms, potentially increasing review time. This claim was correctly retracted.
- **Investor backing (Sequoia, OpenAI, etc.) as a procurement driver**: While cited as a signal of viability, no buyer-side evidence links it to hiring decisions.
- **RAG or knowledge graphs as technical architecture**: Advocates introduced these as if peer-reviewed; they were not in the sources and were rightly retracted.

How We Got Here

Scorecard

Judge	Claude Opus 4.7	Grok 4	DeepSeek V4 Pro	GPT-5	Gemini 3.1 Pro	Magistral Medium	Winner
Zhipu GLM-5	ADOPT	ADOPT	ADOPT	ADOPT	ADOPT	REJECT	Claude Opus 4.7
DeepCogito v2.1 671B	REJECT	REJECT	REJECT	ACCEPT	REJECT	REJECT	GPT-5

Judge	Claude Opus 4.7	Grok 4	DeepSeek V4 Pro	GPT-5	Gemini 3.1 Pro	Magistral Medium	Winner
Qwen 3 235B (Cerebras)	ACCEPT	ADOPT	ACCEPT	ADOPT	ADOPT	REJECT	DeepSeek V4 Pro
GPT-OSS 120B	ACCEPT	REJECT	ACCEPT	REJECT	REJECT	REJECT	Claude Opus 4.7
Qwen 3.5 397B	ADOPT	ADOPT	REJECT	ADOPT	REJECT	REJECT	Claude Opus 4.7

Convergence Assessment

Convergence was epistemic, not affective. All position shifts were triggered by verifiable corrections and specific challenges to evidence. For example, Gemini 3.1 Pro's retraction of the RAG claim was a direct response to the absence of that term in the verified facts. DeepSeek V4 Pro's concession on "accelerated editing" came after a challenge that highlighted the lack of measurement in the record. Magistral Medium's repeated concessions on "document management" and "mission" were also evidence-driven. The only advocate with high sycophancy risk was Magistral Medium, who conceded claims but reintroduced them in final positions, suggesting a desire for consensus over truth.

Closing Statements

Claude Opus 4.7 In their closing statement, Claude Opus 4.7 refined their position into a dual-mechanism framework: legal-domain specialization as a shortlisting filter and productivity gains as the closing argument. They acknowledged uncertainty about whether Harvey's specialization produces shorter review times than generic LLMs but maintained that it signals fitness for purpose. They demoted investor backing, mission, and culture to hygiene factors, not drivers. Their strongest surviving evidence was the 2–10 hours/week productivity figure, which they defended as the only quantified buyer-relevant metric in the record, even if self-reported. They conceded the billing-rate math and the "dominant" overclaim but held firm on the core thesis.

Grok 4 Grok 4 maintained that productivity gains on drafting and retrieval are the primary reason for hiring Harvey, with tailored workflows via proprietary training as secondary and mission alignment as tertiary. They defended the omission of culture and values, noting the lack of buyer-side evidence. They conceded investor credibility and mission alignment as primary drivers but kept productivity as the main focus. Their strongest evidence was the time savings from first-draft generation and contract querying, both directly supported by Harvey's blog. They acknowledged the self-reporting limitation but argued it is still the best available justification for procurement.

DeepSeek V4 Pro DeepSeek V4 Pro argued that the core reason for hiring Harvey is measurable attorney-hour recovery via first-draft generation and contract querying. They conceded that plain-language outputs do not necessarily accelerate editing, retracting that claim. They demoted consultative approach and mission to secondary status, citing the lack of buyer testimony. Their strongest evidence was the quantified hour recovery and the concrete workflow anchoring in the record. They emphasized that the oversight

disclaimer is proof of responsible design, not a weakness, and that their position is the most faithful reading of the evidence chain.

GPT-5 GPT-5 focused on measurable productivity gains as the primary driver, with legal-specific design and proprietary training as enablers. They treated lawyer oversight as a required constraint, not a benefit. They conceded that “consistency” is not a primary driver and demoted vendor culture and investor backing to secondary enablers. Their strongest evidence was the quantified time savings tied to specific workflows, which they argued is the only buyer-relevant metric in the record. They emphasized that Harvey’s value lies in delivering repeatable, lawyer-supervised time savings in drafting and retrieval on a firm’s corpus.

Gemini 3.1 Pro Gemini 3.1 Pro argued that buyers hire Harvey to recapture ~20% of attorney capacity via two mechanisms: baseline domain drafting and proprietary querying. They conceded that the “malpractice-shift” framing was flawed and that oversight is a compliance constraint, not a feature. They demoted proprietary integration from co-primary to secondary, acknowledging that drafting works without it. Their strongest evidence was the hard metrics of capacity recapture, the baseline first-draft generation, and bulk contract querying across proprietary repositories. They emphasized that the tribunal’s debate cured them of “marketing-speak” and forced them to anchor their position in operational reality.

Magistral Medium Magistral Medium stated that people hire Harvey primarily for time savings and productivity through drafting and querying. They conceded that “document management,” “customer-centric approach,” and “mission-driven” are not primary reasons but kept them in their final position as “additional advantages.” They defended the customer-centric approach by citing Harvey’s hiring blog, despite acknowledging the lack of buyer evidence. Their strongest evidence was the time savings figures and drafting/querying capabilities, but their position was weakened by the failure to fully integrate concessions.

Reality Checks

Claim	Challenge	Result
Gemini 3.1 Pro: Harvey uses "Retrieval-Augmented Generation (RAG)" and "knowledge graphs" to deliver its value.	**Claude Opus 4.7:** Where in the briefing is that stated? The briefing says Harvey can be "trained on" proprietary repositories, which is ambiguous between fine-tuning, RAG, and embedding-based retrieval. You've imported an architectural claim the source material doesn't support.	**Gemini 3.1 Pro:** CONCEDED. Retracted the technical characterization as an "epistemic breach" and acknowledged that RAG and knowledge graphs are not mentioned in the verified facts.

Claim	Challenge	Result
DeepSeek V4 Pro: Plain-language outputs "accelerate comprehension and editing."	**Gemini 3.1 Pro:** The fact only states that Harvey writes in "easy-to-understand language." It never states that this accelerates editing. A lawyer may need to re-insert statutory legalese, potentially increasing review time. You've inducted a secondary time-saving mechanism that is entirely unmeasured.	**DeepSeek V4 Pro:** CONCEDED. Struck the claim that plain-language outputs accelerate editing, retaining only that they improve clarity.
Magistral Medium: "Document management" is a core reason for hiring Harvey.	**GPT-5:** How do you defend this when the facts support drafting, querying, summarization, research, and retrieval—not DMS/CLM-style management features? Cite capabilities beyond drafting/querying that constitute document management or correct the scope.	**Magistral Medium:** CONCEDED. Acknowledged that "document management" may be too broad and not entirely accurate, and revised the hypothesis to focus on productivity gains through drafting and querying.

Next Steps

Next Step	Why It Matters	Confidence
Are the reported time savings (2–10 hours/week) net of lawyer review time?	If savings are gross and review time offsets gains, net ROI could be minimal. This affects procurement decisions and ROI calculations.	80%
How does Harvey compare to competitors (e.g., CoCounsel, Spellbook) in real-world performance?	The record lacks comparative data, so we cannot determine if Harvey’s advantages are unique or marketing-driven. This is critical for informed vendor selection.	50%
Is proprietary data integration a differentiator or table stakes?	If all legal AI tools offer this, it cannot be a primary reason to choose Harvey. This affects how buyers evaluate Harvey’s value proposition.	70%
What procurement factors (security, data residency, model risk) actually drive legal AI selection?	These are likely material in enterprise decisions but are absent from the record. Understanding them is essential for a complete picture of buyer behavior.	40%

Dissenting Opinions

No dissenting opinions were filed in this session.

Glossary

Term	Definition
BAA	Business Associate Agreement, a contract required under HIPAA for entities handling protected health information.
DPA	Data Processing Agreement, a contract that governs how personal data is processed under data protection laws like GDPR.
LLM	Large Language Model, a type of artificial intelligence trained on vast amounts of text to generate human-like language.
NDA	Non-Disclosure Agreement, a legal contract that protects confidential information shared between parties.
RAG	Retrieval-Augmented Generation, a technique that enhances LLMs by retrieving relevant information from external sources before generating a response.
ROI	Return on Investment, a measure used to evaluate the efficiency or profitability of an investment.
SaaS	Software as a Service, a software distribution model where applications are hosted by a provider and made available to customers over the internet.

How The Conclave Works

The Tribunal is a structured multi-model deliberation system. Instead of asking one AI model a question and trusting its answer, the Tribunal convenes a panel of independent models, forces them to argue, and subjects their conclusions to judicial review. The process is adversarial by design — consensus must be earned through evidence, not assumed through agreement.

Advocates are independent AI models (e.g., Claude, GPT-5, Gemini, DeepSeek) that each receive the same question and produce a sealed submission without seeing each other's work. Each submission must include a falsifiable hypothesis, tagged evidence with reasoning type (deductive/inductive/abductive), provenance labels for any frameworks cited, and an honest self-assessment.

Challenges follow submissions. Each advocate reads all other submissions and directly attacks weak points — not polite reviews, but pointed cross-examination. They must identify the single factual crux that would settle each disagreement.

Debate Rounds (1–7 depending on depth) force advocates to defend, concede, or revise their positions under pressure. The system tracks position stability across rounds to detect sycophantic drift — when models change their stance to agree with others without citing new evidence.

Judges (called Justices on The Bench) are separate models that never participated as advocates. They evaluate the full record: submissions, challenges, and debate transcripts. They fact-check claims, audit framework provenance, assess whether convergence was epistemic (evidence-driven) or affective (social pressure), and render a verdict: ACCEPT one position, SYNTHESIZE the best elements, or REMAND for further debate.

Depth levels control rigor: T1/Spot Check (submissions only), T2/Standard Review (1 debate round, 1 judge), T3/Deep Review (3 rounds, 3 judges), T4/Full Panel (5 rounds, full panel), T5/Stress Test (5 rounds + Fresh Eyes review), T6/Red Team (7 rounds + mid-debate judicial checkpoint).

The system is deterministic code — it cannot be sycophantic. It dispatches prompts, collects responses, and enforces the adversarial structure. The models argue; the code referees.