Tribunal Session Summary

Session: conclave-20260514-012219-what-are-the-primary-20b7

Depth: T6

Advocates: 6 | Judges: 5 | Cost: $4.4047 | Time: 30m 08s

THE PROMPT: What are the primary reasons people hire Harvey.ai? Discuss the key benefits and common use cases.

THE RULING: Organizations hire Harvey.ai primarily to achieve substantial, self-reported time savings on routine, document-heavy legal workflows—especially drafting, due diligence, and contract corpus Q&A—based on the expectation of efficiency gains, not independently verified outcomes. Competitive signaling is a meaningful secondary driver, particularly among law firms.

The Prompt

What are the primary reasons people hire Harvey.ai? Discuss the key benefits and common use cases.

The Ruling

Organizations hire Harvey.ai primarily to achieve substantial, self-reported time savings on routine, document-heavy legal workflows—especially drafting, due diligence, and contract corpus Q&A—based on the expectation of efficiency gains, not independently verified outcomes. Competitive signaling is a meaningful secondary driver, particularly among law firms.

This answer synthesizes the tribunal's consensus: time savings are the dominant economic motive, but the evidence is entirely vendor-sourced and unverified. The ruling reflects the majority opinion that adoption is driven by the *promise* of efficiency, with market scale (700 clients, $100M+ ARR) signaling procurement readiness, not proven ROI. The key caveat is that no independent studies validate the claimed time savings, net of QA overhead, and savings are variable and task-dependent.

Core Analysis

The Primary Reason: Expected Time Savings on Document-Heavy Workflows

The core reason organizations hire Harvey.ai is the expectation of significant time compression on high-volume, repetitive legal tasks. This expectation is grounded in self-reported data from Harvey and its customers, not independently verified performance. The most frequently cited benefits—time savings, faster cycle times, and integration with existing tools—are consistently tied to three dominant use cases: drafting and markup of legal documents, due diligence in large data rooms, and querying across extensive contract repositories using Harvey Vault.

Harvey markets itself as “Professional Class AI,” a domain-tuned assistant for legal and professional services work. Its value proposition centers on automating labor-intensive tasks that consume attorney hours, such as generating first drafts of NDAs, reviewing key terms in lengthy agreements, and retrieving specific clauses from hundreds of contracts. According to Harvey’s own reporting, internal teams save 20–40 attorney hours per week, while external users report average savings of 2–10 hours per attorney per week, with top-decile users claiming 30–88 hours per month. These figures, however, are **self-reported**, lack third-party validation, and may reflect gross time savings before accounting for quality assurance, rework, or hallucinations.

Key Benefits

The key benefits of using Harvey.ai are derived from its ability to compress time on routine legal work:

- **Time Savings and Productivity Gains**: Users report saving 2–10 hours per week per attorney, with top-decile users saving 30–88 hours per month. These savings allow legal teams to handle more work internally before engaging outside counsel, shortening overall project timelines.
- **Efficiency in Complex Tasks**: Harvey excels at processing large document sets, providing rapid overviews and issue identification that outperform manual reviews. For instance, it enables teams to navigate data rooms "significantly more efficiently than has ever been possible by human effort," freeing up attorneys for strategic work.
- **Scalability and Integration**: As an efficiency tool, Harvey supports work across devices and integrates with familiar systems like Microsoft tools, making it accessible for large firms and global operations. This scalability is key to its adoption in 63 countries, where it helps manage high-volume legal demands without proportional increases in headcount.
- **ROI and Cost Reduction**: By automating routine tasks, Harvey reduces the need for junior attorney hours on drafting or analysis, with benefits like saving "days worth of attorney time" on investigations. This positions it as a high-ROI investment, backed by its $100M+ ARR milestone.

Common Use Cases

Harvey's use cases focus on daily legal workflows, particularly drafting, analysis, and compliance:

- **Drafting and Document Generation**: Generating first drafts of NDAs, contractor agreements, internal policies/guidelines, memos; pleadings/briefs where permitted by practice standards.
- **Contract Review and Markup**: Extract key terms, compare to playbooks, redline counterparties' paper using firm templates in Word.
- **Due Diligence/Data Rooms**: Generate overviews, spot issues, surface key documents rapidly.
- **Knowledge Retrieval**: Query across 1,000+ contracts and internal repositories for specific clauses/obligations.
- **Internal Investigations/Compliance**: Use Harvey Vault to answer complex questions over large document sets.
- **Security Questionnaires**: Draft first-pass responses from a maintained Vault of security policies.

These use cases are indicative of broader adoption, with Harvey serving both law firms and corporate legal departments to streamline operations at scale.

How We Got Here

Scorecard

Judge	Claude Opus 4.7	GPT-5	DeepSeek V4 Pro	Grok 4	Gemini 3.1 Pro	Magistral Medium	Winner
Qwen 3 235B (Cerebras)	ACCEPT	ACCEPT	ADOPT [element]	ADOPT [element]	ADOPT [element]	REJECT	Claude Opus 4.7

Judge	Claude Opus 4.7	GPT-5	DeepSeek V4 Pro	Grok 4	Gemini 3.1 Pro	Magistral Medium	Winner
Qwen 3.5 397B	ACCEPT	REJECT	REJECT	REJECT	REJECT	REJECT	Claude Opus 4.7
GPT-OSS 120B	ACCEPT	REJECT	REJECT	REJECT	REJECT	REJECT	Claude Opus 4.7
DeepCogito v2.1 671B	ACCEPT	REJECT	REJECT	REJECT	REJECT	REJECT	Claude Opus 4.7
Zhipu GLM-5	ADOPT	ADOPT	ADOPT	ADOPT	ADOPT	REJECT	SYNTHESIZE

Convergence Assessment

Convergence was epistemic, not affective. All advocates shifted positions in response to specific evidence challenges—Gemini 3.1 Pro's mathematical errors, Magistral Medium's source independence flaw, DeepSeek V4 Pro's accuracy claim—rather than social pressure. For example, Claude Opus 4.7 revised their use-case rankings after GPT-5 challenged the lack of comparative data, and Gemini 3.1 Pro abandoned their "Efficiency Multiplier" framework after DeepSeek V4 Pro exposed its lack of provenance. These shifts were triggered by verifiable corrections, not sycophantic drift.

Closing Statements

Claude Opus 4.7 My final position is that people hire Harvey primarily for expected attorney-hour compression on document-heavy legal workflows—drafting, due diligence, and contract Q&A—with savings that are variable and task-dependent. Competitive signaling is a meaningful secondary driver, especially for law firms. I learned from Gemini 3.1 Pro that ARR cannot distinguish between genuine ROI and FOMO, forcing me to drop ARR as evidence against pure hype. My strongest surviving evidence is the three documented workflow pillars, which appear in multiple sources, and the fact that the time-savings claim is plausible given LLMs' known capabilities in drafting and synthesis.

DeepSeek V4 Pro My final position is that the primary reason for hiring Harvey is the compression of routine legal work, as consistently described by users and vindicated by the market’s willingness to pay. I learned from Magistral Medium that I must not assert what the evidence cannot support, leading me to remove the "without sacrificing accuracy" claim. My strongest surviving evidence is the triangulation of time-savings reports from multiple internal vantage points, the Hengeler Mueller case study, and market adoption at scale as a revealed-preference test.

Gemini 3.1 Pro My final position is that organizations hire Harvey based on the promise of recovering 2 to 10 hours per week on routine "first-pass" work, with elite trust signals and defensive procurement as key drivers. I learned from Magistral Medium that I was conflating procurement intent with verified ROI, leading me to abandon the "10% to 20% capacity recovery" claim. My strongest surviving evidence is the epistemic void of independent ROI verification, the unmatched power of Harvey's elite trust signals, and the plausibility

of the targeted use cases.

Magistral Medium My final position is that the primary reasons for hiring Harvey are time savings, efficiency gains, and integration with existing tools, though I acknowledge the evidence is self-reported. I learned from GPT-5 and Grok 4dvocate-F that I must be more careful about what conclusions can be drawn from financial metrics. My strongest surviving evidence is the reported time savings, specific use cases from Hengeler Mueller and Harvey's own team, and integration with Microsoft tools.

GPT-5 My final position is that people hire Harvey to compress legal cycle times on document-heavy, repeatable workflows, with drafting, review, and corpus Q&A as the key mechanisms. I learned from Magistral Medium that I must segment drivers by buyer type, leading me to treat Vault-led analysis as a material driver for in-house teams. My strongest surviving evidence is the consistent, material time savings reported by users, the due diligence case study at Hengeler Mueller, and the on-template first-draft generation within Microsoft Word.

Grok 4 My final position is that the primary reasons for hiring Harvey are promised time savings on routine legal tasks, improved efficiency in document review and drafting, and integration with existing tools, though I acknowledge the evidence is self-reported and variable. I learned from Gemini 3.1 Pro that I must not treat vendor blogs as "empirical evidence" of realized outcomes. My strongest surviving evidence is the reported time savings, specific use cases, and integration with Microsoft tools.

Reality Checks

Claim	Challenge	Result
Gemini 3.1 Pro: "Harvey routinely recovers 10% to 20% of an attorney’s billable capacity."	**DeepSeek V4 Pro:** "2 hours per week is only 5% of a 40-hour week; your 10% floor is mathematically impossible."	Full concession; Gemini 3.1 Pro revised the claim to 5–25% and dropped the "Efficiency Multiplier" framework.
DeepSeek V4 Pro: "Harvey enables substitution without sacrificing accuracy."	**GPT-5:** "No independent accuracy benchmarks exist; you're asserting quality without evidence."	Full concession; DeepSeek V4 Pro removed the claim and added a caveat that accuracy is unbenchmarked.
Claude Opus 4.7: "Drafting is the highest-frequency use case."	**GPT-5:** "No usage telemetry exists; you're inferring frequency from marketing emphasis."	Partial concession; Claude Opus 4.7 revised to "prominent, repeatedly-cited" use case.
Magistral Medium: "Consistency across sources validates the time-savings claims."	**Gemini 3.1 Pro:** "The 'sources' are all Harvey's blogs; consistency within a single vendor's marketing is not independent corroboration."	Partial concession; Magistral Medium acknowledged the limitation but continued to cite consistency as a credibility booster.

Next Steps

Next Step	Why It Matters	Confidence
Conduct independent, controlled studies to measure net time savings after accounting for QA and rework.	Gross time savings may be offset by rework if outputs contain hallucinations or errors; net savings determine real ROI.	40%
Determine the proportion of Harvey's $100M ARR that comes from renewals vs. new pilots.	Renewals would indicate sustained value; pilot proliferation suggests experimentation without proven benefit.	30%
Investigate how adoption drivers differ between in-house legal teams and law firms.	In-house teams prioritize cost efficiency; firms face dual pressures of efficiency and competitive signaling; understanding this split improves targeting and evaluation.	60%
Measure Harvey's accuracy rate (e.g., hallucination rate, clause extraction precision) compared to human review or generic LLMs.	Accuracy directly impacts risk and trust; a tool that saves time but introduces legal errors could be counterproductive.	40%

Dissenting Opinions

No dissenting opinions were filed in this session.

Glossary

Term	Definition
ARR	Annual Recurring Revenue, a standard SaaS metric indicating predictable, ongoing revenue from subscriptions.
Due Diligence	The process of investigating and evaluating a business or asset before a transaction, often involving review of large document sets.
FOMO	Fear Of Missing Out, a psychological driver where organizations adopt technology to avoid falling behind peers.
Gross Time Savings	The total time saved by a tool before accounting for time spent on quality assurance, correction, or rework.
Net Time Savings	The actual time saved after accounting for quality assurance, correction, or rework, representing true efficiency gain.
ROI	Return on Investment, a performance measure used to evaluate the efficiency of an investment.
Vault	Harvey's feature for querying and analyzing large sets of documents, such as contracts or policies, using AI.

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.