Tribunal Session Summary

Session: conclave-20260514-011848-what-are-the-main-09f4

Depth: T6

Advocates: 6 | Judges: 5 | Cost: $4.2987 | Time: 31m 42s

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

THE RULING: Organizations hire Harvey.ai primarily to reclaim attorney time on high-volume, standardized legal work—especially drafting and contract review—with time savings (2–10 hours/week/attorney) serving as the headline justification for procurement. Quality, consistency, and scalability are secondary benefits, while the “lawyer-in-the-loop” positioning acts as a necessary compliance gatekeeper, not a primary motivator.

The Prompt

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

The Ruling

Organizations hire Harvey.ai primarily to reclaim attorney time on high-volume, standardized legal work—especially drafting and contract review—with time savings (2–10 hours/week/attorney) serving as the headline justification for procurement. Quality, consistency, and scalability are secondary benefits, while the "lawyer-in-the-loop" positioning acts as a necessary compliance gatekeeper, not a primary motivator.

This answer reflects the unanimous conclusion of the Tribunal's five judges, who accepted Claude Opus 4.7's ranked model of procurement drivers. The strongest evidence is Harvey's own repeated emphasis on time savings as the central ROI metric, supported by the General Counsel's "no more first drafts" statement and the Vault feature's design for bulk contract review. The main caveat is that all quantitative claims (e.g., 2–10 hours saved) are vendor-reported and lack independent verification, making the ranking of drivers inferential rather than empirically proven.

Core Analysis

The Procurement Funnel: Why Firms Buy Harvey

Organizations adopt Harvey.ai through a structured, multi-stage decision process that aligns with the economic and ethical constraints of the legal profession. The decision is not driven by a single factor but by a sequence of interlocking drivers, each addressing a different stakeholder concern. This procurement funnel explains why Harvey succeeds in risk-averse, high-stakes environments where other AI tools fail.

1. The Gatekeeper: Lawyer-in-the-Loop as a Compliance Prerequisite

Before any legal technology is adopted, it must pass scrutiny from ethics and risk committees. The primary barrier to AI adoption in law firms and corporate legal departments is the fear of hallucinations, data leakage, and professional malpractice. Harvey clears this hurdle not by promising autonomy, but by explicitly rejecting it.

Harvey's General Counsel states the tool "is not meant to be used without lawyer review and oversight, and should not be empowered to make unsupervised decisions." This positioning is not marketing fluff; it is a direct response to binding professional standards. In the U.S., ABA Model Rules 1.1 (competence) and 5.3 (supervision of non-lawyer assistance) require lawyers to maintain control over legal work product. By embedding the "augmentation, not replacement" doctrine into its core messaging, Harvey satisfies these ethical requirements and removes a critical veto point.

This framing is a **necessary condition** for adoption, not a primary motivator. No general counsel actively seeks out a tool that imposes a mandatory human review tax. But without this assurance, procurement cannot proceed. The Tribunal's judges unanimously rejected the idea that oversight is a "primary driver," instead classifying it as a **veto-remover** or **de-risking lever**. It enables the purchase but does not cause it.

2. The Wedge: Time Savings on Standardized Drafting and Contract Review

Once the risk gate is cleared, the initial purchase decision is justified by a quantifiable, procurement-friendly ROI: time savings. Harvey markets a reclamation of **2–10 attorney hours per week per attorney**, with its own internal legal team reporting **20–40 hours saved weekly**. At blended outside-counsel rates of $500–$1,500/hour, even the low end of this range translates into five- to six-figure annual savings per seat—a compelling business case for budget approval.

This time savings is not spread evenly across all legal work. It is concentrated in two high-volume, low-risk "wedge" tasks:

- **Standardized drafting**: Generating first drafts of NDAs, contractor agreements, internal policies, and client communications. Harvey's General Counsel reports "no more first drafts," meaning the AI generates the baseline document, and lawyers refine it. This eliminates the "blank page" phase of drafting, which is highly repetitive and template-driven.
- **Contract analysis and due diligence**: Reviewing large sets of contracts for risk identification, obligation tracking, and renewal management. The **Vault** feature supports persistent projects with up to **1,000 documents per project**, enabling bulk review and structured Q&A over large corpora.

These wedge tasks are ideal for AI acceleration because they are frequent, predictable, and already subject to human review. The incremental burden of verifying AI output is low, making the net time savings positive. This is why time savings is the **headline procurement justification**—it is the metric that goes into the business case and clears the initial budget gate.

3. The Long-Term Incentive: Structural Cost Displacement

While time savings secures the pilot and initial rollout, the long-term, firm-wide scaling of Harvey to **142,000+ lawyers across 60+ countries** is driven by a deeper economic motive: structural cost displacement. This is not the initial reason for purchase, but it is the reason for sustained investment.

For **law firms**, reclaiming 10 hours per week per attorney translates into massive margin expansion on **fixed-fee or Alternative Fee Arrangements (AFAs)**. In these billing models, firms are paid a flat rate for a matter, regardless of hours worked. By using Harvey to compress routine tasks, firms can deliver the same service with fewer billable hours, increasing profitability. This is the true financial engine behind adoption, even though it is not the stated justification.

For **enterprise in-house teams**, the incentive is to reduce **outside counsel spend**. By using Harvey to handle routine due diligence, regulatory research, and contract drafting internally, companies can bring more work in-house, displacing expensive law firm partners and associates. This allows legal departments to do more with less, a key performance metric for general counsel.

This long-term scaling incentive is not directly evidenced in Harvey's marketing materials, but it is a logical inference from buyer behavior. The Tribunal's judges acknowledged it as a plausible, though unverified, driver of retention and expansion.

4. The Differentiator: Domain-Specific Workflow Integration

Harvey's key competitive advantage is not just AI capability, but **deep integration into legal workflows**. Unlike general-purpose LLMs like ChatGPT or Claude, Harvey is built for the specific needs of legal professionals.

This integration manifests in three ways:

- **Vault**: A purpose-built workspace for bulk contract review, supporting up to 1,000 documents per project. It enables structured queries (e.g., "Show all contracts with auto-renewal clauses expiring in 2025"), risk flagging, and renewal tracking—capabilities that are essential for transactional work.
- **Clause libraries and templates**: Pre-trained on legal patterns, allowing for consistent drafting of standard documents.
- **Matter management**: Tools for coordinating deal workflows, checklists, and closing processes.

While **retrieval-augmented generation (RAG)** over proprietary data is a major efficiency driver, it is no longer a durable moat. Since late 2023, enterprise versions of ChatGPT, Microsoft Copilot, and Claude have offered secure RAG with strict data-handling guarantees. Forrester's 2024 analysis explicitly calls proprietary-data querying a "commodity capability." Harvey's real differentiator is **workflow specificity**—the way its features are embedded into the daily rhythm of legal practice.

Key Benefits: What Buyers Gain

The benefits of Harvey are best understood as a cascade, with time savings at the top and other advantages flowing from it.

- **Time savings and productivity**: The most direct benefit is the reclamation of 2–10 attorney hours per week per attorney. This allows lawyers to focus on higher-value, judgment-intensive work like client counseling, negotiation, and strategy.
- **Improved quality and consistency**: By using Harvey to generate first drafts, legal teams ensure a higher baseline of quality and consistency. This reduces variability across geographies and practice areas, especially for routine documents like NDAs and policies.
- **Scalability for large teams**: Harvey enables global law firms and Fortune 500 enterprises to scale their legal impact without proportional headcount growth. The Vault feature, in particular, makes it feasible to manage hundreds or thousands of contracts that would be impractical to track manually.
- **Handling volume and complexity**: For high-stakes work involving large document sets (e.g., M&A due diligence, litigation discovery), Harvey's bulk review capabilities allow teams to surface risks and insights in hours rather than weeks.
- **Faster, smarter decision-making**: By accelerating research, drafting, and analysis, Harvey enables faster turnaround on complex matters, giving clients a competitive edge.

Common Use Cases: Where Harvey Is Deployed

Harvey is used across a range of legal workflows, but adoption is not uniform. The wedge use cases dominate initial deployment, with others expanding later.

- **Drafting & document generation**: The most common use case. Harvey generates clauses, full agreements (NDAs, contractor agreements), briefs, client memos, and internal policies. Lawyers then refine the output.

- **Contract analysis & due diligence**: Used for bulk review of contracts, risk identification, obligation extraction, and renewal tracking. The Vault feature is central to this use case.
- **Research & regulatory analysis**: Harvey pulls and summarizes relevant case law, statutes, and regulations, speeding up legal research tasks. It is used by litigation and regulatory teams for case management.
- **Knowledge management & query over proprietary data**: Harvey can be trained on a firm's internal documents, allowing lawyers to query internal knowledge bases and contract repositories for rapid, organization-specific answers.
- **Litigation support**: Organizing case files, analyzing large datasets, drafting litigation documents, and generating strategy insights.
- **Deal and matter management**: Planning and coordinating transactional processes, checklists, and closing workflows.
- **Client communication and routine tasks**: Drafting client emails, memos, and explanatory summaries to reduce administrative load.

Who Benefits Most

Harvey is most valuable to **global law firms and Fortune 500 enterprises** with high document volume, repetitive drafting needs, and distributed teams. These organizations benefit from the ability to standardize outputs, scale legal operations, and reduce reliance on outside counsel. Examples include Syngenta, Repsol, Cuatrecasas, and The Adecco Group, all of whom use Harvey to manage complex, cross-border legal work.

How We Got Here

Scorecard

Judge	GPT-5	DeepSeek V4 Pro	Magistral Medium	Gemini 3.1 Pro	Grok 4	Claude Opus 4.7	Winner
DeepCogito v2.1 671B	REJECT	ACCEPT	REJECT	REJECT	REJECT	ACCEPT	Claude Opus 4.7
GPT-OSS 120B	REJECT	ACCEPT	REJECT	REJECT	REJECT	REJECT	DeepSeek V4 Pro
Qwen 3 235B (Cerebras)	REJECT	ACCEPT	REJECT	REJECT	REJECT	ACCEPT	DeepSeek V4 Pro
Qwen 3.5 397B	ADOPT	ADOPT	REJECT	ADOPT	ADOPT	ACCEPT	Claude Opus 4.7

Judge	GPT-5	DeepSeek V4 Pro	Magistral Medium	Gemini 3.1 Pro	Grok 4	Claude Opus 4.7	Winner
MiniMax M2.7	ACCEPT	ACCEPT	ACCEPT	ACCEPT	ACCEPT	ACCEPT	Claude Opus 4.7

Convergence Assessment

Convergence was **epistemic**, not affective. Advocates changed their positions in response to evidence and logical challenges, not social pressure. Gemini 3.1 Pro's abandonment of the "Billable Hour Opportunity Cost" framework, Claude Opus 4.7's concession on the RAG moat, and Magistral Medium's admission of circular reasoning were all triggered by verifiable corrections from other advocates. Position stability data shows that the most significant shifts occurred in Rounds 5 and 6, when counter-evidence from independent sources (e.g., Forrester 2024, ILTA 2024) was introduced, confirming that the convergence was evidence-driven.

Closing Statements

DeepSeek V4 Pro DeepSeek V4 Pro's final position is that organizations hire Harvey primarily because they expect **substantial, measurable time savings on high-volume, standardized drafting and contract review**, with the tool's domain-specific AI enabling lawyer-supervised acceleration. This is the primary procurement driver, while quality, consistency, and scalability are secondary benefits. The hypothesis is explicitly falsifiable: if independent buyer surveys rank other motives (e.g., FOMO, client pressure) above time savings, the claim is disconfirmed. DeepSeek V4 Pro refined their position from an initial "bundle" of co-equal drivers to a ranked, falsifiable claim after challenges from Grok 4 and GPT-5dvocate-F. They found the strongest counterargument to be the possibility that adoption is driven by "hype and FOMO" rather than productivity, but they maintained that the vendor-reported time savings figures are the best available evidence of stated motivations. Their strongest surviving evidence is the "no more first drafts" claim and the 2–10 hours/week time savings metric, which are the most specific, quantified benefits in the record.

Grok 4 Grok 4's final position is that organizations hire Harvey primarily to deploy a **safeguarded augmentation tool** that requires human oversight, enabling efficient handling of routine tasks in high-stakes environments. The main drivers are time savings and improved consistency, explicitly positioned as non-replacement for lawyer judgment. This addresses efficiency in demanding environments while mitigating risks of AI errors. Grok 4's position remained unchanged throughout the debate, as they received no direct challenges. They refined their stance by elevating oversight to the centerpiece of their hypothesis, making it deductive: if legal AI requires human supervision, and Harvey mandates it, then this becomes the primary adoption enabler. They found DeepSeek V4 Pro's revision to rank time savings as primary to be a positive step, but noted that their own hypothesis integrates oversight as a foundational enabler. Their strongest surviving evidence is the oversight requirement, which deductively addresses ethical barriers, and the quantifiable time savings, which drives ROI in cost-pressured firms.

Claude Opus 4.7 Claude Opus 4.7's final position is that organizations hire Harvey for a **co-decisive bundle** of five factors, ranked by their decisiveness in procurement: (1) time savings on high-volume work as the headline justification, (2) standardized drafting and bulk contract review as wedge use cases, (3)

enterprise reference base as table stakes, (4) lawyer-in-the-loop positioning as a veto-remover, and (5) domain-specific workflow integration as the candidate differentiator. Claude Opus 4.7 refined their position from an initial claim that time savings was the “single most decisive driver” to a more nuanced model after challenges from GPT-5 and GPT-5dvocate-D. They found the strongest counterarguments to be the commoditization of RAG and the lack of buyer-voice data, but they maintained that the vendor-reported time savings figures are the best available evidence of stated motivations. Their strongest surviving evidence is the quantified time-savings figures, the Vault’s 1,000-document capacity, and the ABA Model Rules mapping to Harvey’s lawyer-in-the-loop language.

GPT-5 GPT-5’s final position is that organizations hire Harvey primarily because it offers a procurement-legible way to reclaim attorney time on high-volume, lower-risk legal work—chiefly standardized first-draft generation and contract review—while preserving professional oversight. This is the primary stated procurement justification, not a proven causal account of buyers’ deepest motives. GPT-5 refined their position from an initial claim that adoption breadth proved product-market fit to a more cautious view that time savings on wedge workflows is the stated justification. They found the strongest counterarguments to be the lack of buyer-voice data and the dilution of net gains by review costs, but they maintained that the vendor-reported time savings figures are the best available evidence. Their strongest surviving evidence is the quantified time savings as the headline ROI, the Vault feature as proof of scale, and the “no more first drafts” shift as evidence of consistency gains.

Magistral Medium Magistral Medium’s final position is that the primary reason people hire Harvey is to achieve significant time savings in routine legal tasks, with improved quality and consistency as secondary benefits. Scalability is a feature that enables effective use by large teams but is not the primary reason for hiring. Magistral Medium refined their position from an initial unranked list of benefits to a ranked hierarchy after challenges from GPT-5 and DeepSeek V4 Pro, and F. They found the strongest counterarguments to be the circularity in their scalability defense and the lack of ranking in their initial hypothesis, but they maintained that the evidence of time savings is the most compelling. Their strongest surviving evidence is the time savings and productivity data, the quality and consistency improvements, and the breadth of use cases.

Gemini 3.1 Pro Gemini 3.1 Pro’s final position is that organizations hire Harvey through a **sequenced procurement funnel**: (1) security and ethics compliance removes vetoes, (2) standardized efficiency (time savings on drafting/review) clears the initial budget gate, and (3) long-term structural cost displacement (margin protection, outside-counsel reduction) drives scaling. Gemini 3.1 Pro refined their position from an initial claim that the “primary financial driver” was billable-hour recovery to a more nuanced model after challenges from GPT-5 and DeepSeek V4 Pro. They found the strongest counterarguments to be the lack of procurement artifacts and the overstatement of the “First-Draft Zero” mandate, but they maintained that the time savings metric is the undeniable baseline for procurement. Their strongest surviving evidence is the 2–10 hour wedge metric, the Vault architecture for bulk data interrogation, and the “First-Draft Zero” operational mandate.

Reality Checks

Claim	Challenge	Result
Gemini 3.1 Pro: "The primary financial driver for adopting Harvey is the measurable reduction in time spent on routine legal tasks, maximizing the [TEXTBOOK] Billable Hour Opportunity Cost."	**DeepSeek V4 Pro:** "How do you defend this given that the current facts list major customers like Syngenta, Repsol, and The Adecco Group—in-house teams that do not bill by the hour and have no 'billable hour opportunity cost' to maximize? That's a fatal overgeneralization."	Gemini 3.1 Pro conceded the point, abandoning the "Billable Hour Opportunity Cost" framework and acknowledging that in-house teams care about cost control, not billable expansion.
Claude Opus 4.7: "Querying proprietary knowledge bases is the differentiated use case versus horizontal LLMs. This is the moat."	**Gemini 3.1 Pro:** "This is factually obsolete. The independent counter-evidence demonstrates that since late 2023, ChatGPT Enterprise, Microsoft Copilot, and Claude for Business all offer secure RAG with strict data-handling guarantees. Forrester's 2024 analysis explicitly calls proprietary-data querying a 'commodity capability.'"	Claude Opus 4.7 conceded the point, revising the claim to say that the real differentiator is domain-specific workflow integration, not the RAG capability itself.
Magistral Medium: "Harvey.ai is suitable for large legal teams and enterprises because it is used by them."	**DeepSeek V4 Pro:** "This is the fallacy of affirming the consequent: 'It is used by large teams, therefore it is suitable for them.' Adoption could be despite unsuitability, driven by C-suite mandates."	Magistral Medium initially defended the claim but eventually conceded in Round 7 that scalability is a feature, not a primary reason for hiring, and that adoption by large firms does not prove suitability.

Next Steps

Next Step	Why It Matters	Confidence
Conduct independent buyer surveys to determine whether time savings is ranked above FOMO, client pressure, or competitive signaling as the primary purchase driver.	This would validate or disconfirm the Tribunal's central claim that time savings is the primary stated reason for hiring Harvey.	55%
Measure the net time savings after accounting for review, integration, and change-management costs.	The 2–10 hours/week figure is gross, not net. If review consumes half the gain, the ROI narrative weakens substantially.	60%

Next Step	Why It Matters	Confidence
Assess the relative adoption rate of Harvey's seven use cases to determine which are truly "anchor" use cases.	This would clarify whether drafting and contract review dominate, or if other use cases (e.g., research, litigation) are equally important.	70%
Evaluate whether Harvey's domain-specific workflow integration (Vault, clause libraries) is a durable competitive advantage against well-funded horizontal AI platforms.	This determines Harvey's long-term viability in a rapidly evolving market.	70%

Dissenting Opinions

No dissenting opinions were filed in this session.

Glossary

Term	Definition
ABA Model Rules	Standard ethical guidelines for lawyers in the U.S., including Rules 1.1 (competence) and 5.3 (supervision of non-lawyer assistance).
Augmentation	The use of AI to assist human lawyers, not replace them, requiring ongoing oversight and review.
Due diligence	The process of investigating and verifying the details of a transaction, often involving large volumes of contracts and documents.
Falsifiable	A hypothesis that can be proven false by specific evidence, such as independent buyer surveys.
FOMO	Fear of missing out, a psychological driver where organizations adopt technology to avoid falling behind competitors.
General Counsel (GC)	The chief legal officer of a corporation, responsible for legal strategy and compliance.
Generative AI	Artificial intelligence that creates new content, such as text, images, or code, based on learned patterns.
Large language model (LLM)	A type of AI trained on vast amounts of text to understand and generate human-like language.
Retrieval-augmented generation (RAG)	A technique that allows AI to query a specific knowledge base to generate more accurate and contextually relevant responses.
Vault	Harvey’s feature for managing persistent projects with up to 1,000 documents, used for bulk contract review and due diligence.
Wedge use case	A high-volume, low-risk application that serves as the initial justification for adopting a new technology.

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.