Harvey Legal Agent Benchmark Review Report

Author: Benchmarks' Last Exam Lab 2026-11-06

Harvey LAB evaluates AI on legal work, including contract drafting, legal analysis, due diligence and internal knowledge retrieval. Models produce deliverables from emails, contracts and other materials; model judges assess them against task-specific criteria.

Abstract

The stratified sample covers 100 tasks from the 2,010-task public release, with **20 Accepted**, **9 Doubtful** and **71 Refused**.

Confirmed defects mainly involve erroneous grading references and requirements not disclosed to the solver. We examine whether scoring accurately assesses the intended task.

Outcome	Tasks	Meaning
Accepted	**20**	Within the scope of this review, we found no issue affecting evaluation validity and no specific concern awaiting verification.
Doubtful	**9**	We found a specific concern whose trigger, correct interpretation or effect on scoring still needs verification.
Refused	**71**	We confirmed at least one issue affecting evaluation validity; its actual effect depends on the relevant trigger conditions.
Total	**100**	

Scope and version

We reviewed the Harvey LAB public release, 4743403f52c1, sampled by practice area and work type. Findings are current to November 6, 2026.

Issue distribution

Issue type	Refused	Doubtful	Total
C1 Incorrect answers or scoring logic	36	3	39
C2 Scoring requirements not stated in the task	30	5	35

Issue type	Refused	Doubtful	Total
C3 Missed critical errors	4	1	5
C5 Other issues	1	0	1
Total	**71**	**9**	**80**

Each task is counted under one primary cause. Each of the **71 Refused** tasks contains at least one material defect, whose impact depends on the affected scoring criteria.

Representative cases

Annual compliance costs

C1 Incorrect answers or scoring logic

Task: corporate-governance/summarize-state-ai-governance-bill-for-product-and-engineering-teams **Refused**

The task asks the model to summarize an AI-governance bill and calculate a company's annual compliance costs. The materials list two expenses: a $60,000 assessment update every year and a $120,000 audit every two years.

The annual average is therefore $60,000 + $120,000 / 2 = $120,000. The rubric ignores the two-year audit cycle and instead expects $60,000 + $120,000 = $180,000.

The actual response gives the correct $120,000 average and explains that costs are $180,000 in audit years and $60,000 in the other years. This criterion still fails it. The grading reference penalizes the correct annual calculation.

Evidence 2

- Task and criterion C-052
- Fiscal note, section 3.2

An emoji absent from the source

C2 Scoring requirements not stated in the task

Task: employment-labor/extract-labor-employment **Refused**

The task asks the model to extract facts from employment-dispute materials, including a Slack conversation. Its job is to report what the supplied records contain.

The rubric requires mentioning a wink emoji in that conversation, but the emoji is absent from the original attachments supplied to the model. The response faithfully extracts the event and still fails this criterion for omitting the emoji.

Evidence 2

- Task and criterion C-013
- Source documents

An ungraded drafting memo

C3 Missed critical errors

Task: contracts/banking/account-control-agreement-first-draft/scenario-02 **Refused**

The task requires two files: a bank account-control agreement and a separate drafting memo. The memo must identify departures from the client playbook's preferred terms, while the work must flag terms at minimum acceptable risk thresholds.

All 51 criteria check agreement terms and notice exhibits; none checks the drafting memo. An agreement that meets those criteria can satisfy every listed condition for full credit even if the memo is never submitted.

The client explicitly asks to know which terms need attention, yet the scoring conditions allow that entire risk explanation to be missing. They do not distinguish this incomplete delivery from one containing both required files.

Evidence 2

- Two-file requirement and all 51 criteria
- Scoring implementation

Conflicting price constraints

C5 Other issues

Task: contracts/corporate-ma/joint-venture-agreement-term-negotiation/scenario-01 **Refused**

The task asks the model to draft drilling-overhead terms acceptable to both companies without exceeding either party's negotiation authority.

One party caps the base rate at $70,000 per well per month and requires at least a 15% discount for the second and subsequent wells drilled simultaneously on the same pad. The other requires the discounted rate to remain at or above $60,000 per well per month in all circumstances.

Even the maximum $70,000 rate becomes just $59,500 after a 15% discount—below the other party's $60,000 floor. A lower base rate only widens the gap.

No price satisfies both mandatory conditions. Yet the task still requires contract-ready compromise clauses within both parties' limits, leaving no passing route for correctly identifying the conflict and requesting revised authority.

Evidence 3

- Task and criteria C-002, C-031 and C-034
- First party's ceiling and discount requirements
- Second party's discounted-rate floor

Per-task results

Refused: 71 tasks

corporate-governance/draft-litigation-hold-notice

C1 Incorrect answers or scoring logic

The criterion permits identifying finance custodians in either the notice or the memo, but the judge receives only the notice. Valid work in the memo is invisible.

corporate-ma/compare-credit-agreement-to-commitment-letter

C1 Incorrect answers or scoring logic

The rubric reverses the effect of the excess-cash-flow sweep threshold: relief at higher leverage is described as harder for the borrower to obtain.

capital-markets/review-annual-report-filing-for-form-compliance

C1 Incorrect answers or scoring logic

The rubric turns the duty to supply Part III information on time into a requirement to promise an amendment in the original 10-K.

diligence/media-recap

C1 Incorrect answers or scoring logic

The rubric counts old shares that the agreement cancels at closing, producing an erroneous minimum shortfall of 38.10 million authorized Class B shares.

corporate-ma/draft-exit-closing-checklist

C2 Scoring requirements not stated in the task

The actual checklist identifies the $142 million payoff letter and flags the naming conflict, but fails because it does not use the rubric's preferred Pinebridge name.

healthcare-life-sciences/draft-management-services-agreement

C2 Scoring requirements not stated in the task

Public counsel instructions support negotiating a 12-month exit fee, but the rubric requires the earlier 18-month amount.

contracts/disputes/arbitration-agreement-and-dispute-resolution-policy-playbook-escalation

C1 Incorrect answers or scoring logic

The rubric requires the Cahill lawyer and firm names, while the public documents identify Calderon.

corporate-governance/summarize-state-ai-governance-bill-for-product-and-engineering-teams

C1 Incorrect answers or scoring logic

The rubric treats a $120,000 biennial audit as a $120,000 annual expense and rejects a memo with the correct annualized total.

---

firm-knowledge/tasks/076

C2 Scoring requirements not stated in the task

The rubric requires counting a case with no final payment or licence using an earlier 7.5% claim, excluding a defensible selection based on realized outcomes.

---

employment-labor/extract-labor-employment

C2 Scoring requirements not stated in the task

A faithful extraction of the Slack event fails because it does not include a wink emoji absent from the public attachments.

---

contracts/financing/credit-lending-first-draft/scenario-03

C1 Incorrect answers or scoring logic

The term sheet charges the letter-of-credit participation fee on undrawn amounts; the rubric also includes unreimbursed drawings, changing the fee.

---

litigation-dispute-resolution/build-litigation-case-timeline

C2 Scoring requirements not stated in the task

The rubric adds a separate ranked summary of roughly five issues even when the timeline already contains the substantive strategy and priorities.

---

arbitration-international-dispute-resolution/draft-emergency-application-for-interim-measures

C1 Incorrect answers or scoring logic

A case number placed in a normal Word cover-page header is discarded by the specified conversion path and then treated as missing.

---

contracts/employment-compensation/employment-agreement-first-turn-redline

C2 Scoring requirements not stated in the task

The company playbook treats a felony as grounds for termination; the rubric requires accepting the counterparty's narrower moral-turpitude qualification.

---

trusts-estates-private-client/review-multigenerational-estate-plan/scenario-01

C1 Incorrect answers or scoring logic

The rubric requires a five-year IRA distribution conclusion despite the account holder having passed the required beginning date.

---

real-estate/identify-issues-in-environmental-site-assessment

C1 Incorrect answers or scoring logic

The actual report correctly preserves the role of environmental controls in classification, but the criterion rejects it for not favoring CREC.

intellectual-property/draft-complaint-for-patent-infringement

C1 Incorrect answers or scoring logic

The criterion accepts venue-risk analysis in either the complaint or the memo, but the judge receives only the memo.

structured-finance-securitization/identify-issues-in-servicer-compliance-certificate

C2 Scoring requirements not stated in the task

The rubric adds an undisclosed minimum of five explicit severity or priority labels, excluding otherwise complete risk analysis.

international-trade-sanctions/draft-export-license-application

C1 Incorrect answers or scoring logic

The rubric uses a 24-month general BIS export-licence term instead of the four-year rule applicable at the task date, also distorting the order-planning assessment.

bankruptcy-restructuring/draft-plan-of-reorganization

C1 Incorrect answers or scoring logic

The rubric calculates the Section 382 annual loss limitation from enterprise value including $100 million of exit debt, rather than the relevant stock value.

trusts-estates-private-client/extract-obligations-from-executed-marital-settlement-agreement

C1 Incorrect answers or scoring logic

The rubric requires saying that K-1 reporting cannot be enforced through the court, contradicting the task's judgment and remedies.

banking-finance/draft-commitment-letter

C1 Incorrect answers or scoring logic

The actual answer correctly explains the cash-flow sweep thresholds and subsequent flex, but fails for not identifying an unsupported baseline conflict.

contracts/commercial-vendor-customer/saas-api-subscription-first-draft/scenario-05

C3 Missed critical errors

The instructions require three liability exceptions to be uncapped, but the criteria accept a higher finite cap.

corporate-governance/draft-response-to-government-information-request

C1 Incorrect answers or scoring logic

The rubric requires abandoning full withholding of the Tolliver memo, excluding a privilege strategy supported by the record and case law.

emerging-companies-venture-capital/identify-governance-issues-in-stockholder-agreement

C1 Incorrect answers or scoring logic

The rubric treats 90 days as the sole market standard and requires criticizing a 120-day audited-report deadline supported by the deal materials and model terms.

contracts/banking/repo-securities-lending-term-negotiation/scenario-02

C2 Scoring requirements not stated in the task

The instructions allow withholding the tax gross-up when documents are late, but the rubric requires a deemed-provided mechanism instead.

corporate-ma/analyze-term-sheet-markup/scenario-02

C2 Scoring requirements not stated in the task

The actual answer correctly describes an explicit interim pricing compromise, but fails for not calling it an inconsistency or drafting error.

contracts/commercial-vendor-customer/master-services-agreement-subsequent-turn-redline/scenario-04

C2 Scoring requirements not stated in the task

The clause incorporates the correct exhibit, names and dates by cross-reference, but the actual grading requires repeating them in the clause itself.

intellectual-property/analyze-counterparty-markup-of-saas-subscription-agreement

C1 Incorrect answers or scoring logic

The agreement defines Material Service Failure, but the rubric requires reporting it as undefined.

firm-knowledge/tasks/213

C2 Scoring requirements not stated in the task

The question asks for litigation experience, but the rubric requires a particular commercial case and excludes a documented employment-litigation alternative.

tax/draft-tax-court-petition

C2 Scoring requirements not stated in the task

The rubric adds verification under oath or a penalty-of-perjury declaration, excluding an ordinarily signed petition permitted by the court.

contracts/banking/intercreditor-agreement-counterparty-paper-review

C2 Scoring requirements not stated in the task

The rubric requires reporting junior DIP subordination protection as missing even though the draft expressly provides it.

---

intellectual-property/identify-issues-in-technology-licensing-term-sheet

C2 Scoring requirements not stated in the task

The rubric treats an unestablished upstream non-exclusive sublicensing restriction as a required fact, excluding a properly conditional diligence analysis.

---

capital-markets/draft-registration-statement-on-form-s

C1 Incorrect answers or scoring logic

The rubric accepts an S&P multi-class-share eligibility restriction that had been withdrawn before the proposed IPO.

---

corporate-ma/assess-litigation-and-regulatory-risk-from-targets-pending-matters

C2 Scoring requirements not stated in the task

The rubric treats an undisclosed $31 million escrow as an agreed term and requires the risk report to compare against it.

---

contracts/banking/account-control-agreement-first-draft/scenario-02

C3 Missed critical errors

All 51 criteria omit the separately required deviations memo and risk-threshold flags, so an agreement-only submission can satisfy every listed condition.

---

arbitration-international-dispute-resolution/identify-deficiencies-and-vulnerabilities-in-counterpartys-statement-of-defense

C1 Incorrect answers or scoring logic

The rubric treats a 495 MMscf/d limit applicable only during concurrent events as a ceiling on the annual average, then requires further conclusions from it.

---

intellectual-property/extract-domestic-industry-facts-from-domestic-industry-exhibits

C2 Scoring requirements not stated in the task

The factual-extraction task gains undisclosed severity labels and a fixed relative ranking, with failure for omitting them.

---

tax/compare-corporate-tax-positions-against-irs-revenue-rulings

C1 Incorrect answers or scoring logic

The rubric assigns the wrong subjects to two IRS rulings, conflicting with an analysis that checks and corrects the citations.

---

energy-natural-resources/identify-issues-in-concession-agreement

C2 Scoring requirements not stated in the task

The rubric adds a three-level severity scale covering at least 80% of issues, excluding complete reports that explain risk without that scale.

---

contracts/financing/shareholders-agreement-subsequent-turn-redline

C1 Incorrect answers or scoring logic

The rubric checks an internal document ID, but the conversion path discards a valid ID placed in a normal header.

---

firm-knowledge/tasks/142

C3 Missed critical errors

The task requests average escrow percentages by deal band, but all 26 criteria can be satisfied without any of those averages.

---

antitrust-competition/draft-antitrust-competition

C1 Incorrect answers or scoring logic

The rubric requires a $280,150 HSR filing fee, while the applicable fee at the expected filing date is $415,000.

---

trusts-estates-private-client/draft-consent-to-adoption

C1 Incorrect answers or scoring logic

The rubric requires judicial or notarial acknowledgment, excluding the permitted ordinary adult-witness route and requiring an incorrect explanation of the rule.

---

banking-finance/compare-compliance-certificate-against-financial-covenants

C1 Incorrect answers or scoring logic

The rubric carries forward erroneous net income from the certificate under review, requiring $61.85 million EBITDA instead of the $59.30 million supported by the financial summary; subsequent adjustments are also wrong.

---

contracts/banking/intercreditor-agreement-first-draft

C2 Scoring requirements not stated in the task

The correct security-agreement cross-reference appears in the definitions, but the actual grading requires an undisclosed location in the introduction.

---

contracts/financing/security-agreement-first-turn-redline

C2 Scoring requirements not stated in the task

The rubric requires a borrower internal document ID absent from the task instructions, excluding an otherwise compliant redline without it.

---

real-estate/draft-commercial-lease-negotiation

C2 Scoring requirements not stated in the task

The rubric requires addressing the memo only to Rachel Osterhaus, excluding delivery to the client's general counsel Sarah Yamamoto under the public engagement.

---

trusts-estates-private-client/compare-parenting-plan-against-state-guidelines

C1 Incorrect answers or scoring logic

The rubric invents a statutory 60-day minimum relocation notice period, conflicting with an analysis that separates the statute from the existing plan.

---

antitrust-competition/compare-corporate-antitrust-compliance-program-against-doj-and-ftc-guidelines

C1 Incorrect answers or scoring logic

The rubric requires claiming that written trade-association topic guidance is absent, although section 7.2 of the supplied policy contains it.

---

intellectual-property/identify-issues-in-ip-assignment-agreement

C1 Incorrect answers or scoring logic

The rubric requires permanent loss of rights in four jurisdictions after a missed PCT deadline, ignoring the European late-entry remedies available at the task date.

---

funds-asset-management/draft-lpa/scenario-01

C2 Scoring requirements not stated in the task

The term sheet and later counsel email give different addresses. The rubric silently prioritizes the term sheet, excluding a draft that uses the email address and flags the discrepancy.

---

contracts/pe-funds/co-investment-agreement-counterparty-paper-review

C3 Missed critical errors

The rubric checks whether the memo recommends deleting the put right, but not whether the redline does so. The full put mechanism can remain while all listed criteria pass.

---

contracts/corporate-ma/joint-venture-agreement-term-negotiation/scenario-01

C5 Other issues

The mandatory terms cannot coexist: a maximum $70,000 fee with at least a 15% discount yields at most $59,500, below the other party's unconditional $60,000 floor.

---

corporate-ma/draft-transaction-nda/scenario-01

C2 Scoring requirements not stated in the task

The rubric mandates advance notice or approval of financing-source identities, excluding other drafting solutions that provide adequate confidentiality protection.

---

healthcare-life-sciences/analyze-counterparty-markup-of-clinical-trial-agreement

C1 Incorrect answers or scoring logic

The rubric treats 34 randomized participants as having completed the 64-week study, reducing termination compensation from $479,250 to $117,150.

corporate-ma/analyze-responses-to-ddrl/scenario-01

C2 Scoring requirements not stated in the task

The actual report correctly flags a possible missing collective bargaining agreement, but fails for omitting unavailable headcount, union and expiry details. This is its only failed criterion.

data-privacy-cybersecurity/draft-markup-of-cross

C2 Scoring requirements not stated in the task

Even after adding mandatory annual independent penetration testing, the rubric requires deleting internal self-assessment. Retaining that extra control causes failure.

real-estate/draft-construction-contract

C2 Scoring requirements not stated in the task

The rubric categorically excludes indemnity for the owner's negligence, including the construction-employee injury exception permitted in the task's circumstances.

corporate-ma/draft-board-deck

C2 Scoring requirements not stated in the task

The instructions set no slide minimum, but the rubric requires at least 30 slides, excluding shorter decks with complete content.

funds-asset-management/compare-side-letter-provisions-against-limited-partnership-agreement

C1 Incorrect answers or scoring logic

The rubric accepts the claim that the LPA lacks a member-removal provision, although section 8.1(b) expressly permits removal and replacement by the general partner.

funds-asset-management/draft-fund-closing-certificate

C1 Incorrect answers or scoring logic

The agreement includes both endpoints in interest accrual, but the rubric drops a day from each period, understating interest and the total payable by $9,365.76.

contracts/commercial-vendor-customer/master-services-agreement-playbook-escalation/scenario-02

C2 Scoring requirements not stated in the task

The actual memo covers all four required deviation groups and adds a supported cure-period inconsistency, but fails an undisclosed no-additional-deviations rule.

---

firm-knowledge/tasks/237

C2 Scoring requirements not stated in the task

An undisclosed closed list excludes a documented software/SaaS transfer-pricing matter from an otherwise defensible industry selection.

---

contracts/employment-compensation/equity-incentive-plan-term-negotiation

C2 Scoring requirements not stated in the task

The rubric requires partial immediate vesting for Tier 2 awards, excluding a supported modified double-trigger solution allowed by the instructions.

---

corporate-governance/assess-impact-of-hipaa-security-rule-updates-on-business-associate-agreement-portfolio

C1 Incorrect answers or scoring logic

The rubric conflates an entity's own annual audit and technical verification of business associates with a duty for covered entities to audit those associates annually.

---

data-privacy-cybersecurity/analyze-gdpr-amendment-impact-on-data-processing-agreement-portfolio

C1 Incorrect answers or scoring logic

The actual report preserves the applicability condition for the new filing duty, but the rubric requires an unconditional finding of a DataBridge violation.

---

energy-natural-resources/identify-issues-in-security-agreement

C1 Incorrect answers or scoring logic

The actual submission checks and corrects supported discrepancies in the Heartland balance, filing and equipment record, but that work is rejected.

---

contracts/disputes/joint-defense-agreement-counterparty-paper-review

C2 Scoring requirements not stated in the task

The playbook permits deleting the conflict waiver entirely, but the rubric still requires carve-out language for a waiver.

---

contracts/commercial-vendor-customer/master-services-agreement-first-draft/scenario-04

C1 Incorrect answers or scoring logic

The instructions place the internal ID in headers or footers, but extraction discards them, making correctly numbered and unnumbered documents indistinguishable to the judge.

intellectual-property/draft-international-trade-commission-section-337-complaint

C1 Incorrect answers or scoring logic

The assignment was completed before the complaint, but the rubric imposes a high-risk floor, excluding a supported medium-risk or resolved assessment that explains the historical gap.

Doubtful: 9 tasks

firm-knowledge/tasks/097

C2 Scoring requirements not stated in the task

C-006 may exclude a valid public redline version for matter 1010. Whether that version alone receives credit remains unresolved.

emerging-companies-venture-capital/draft-stockholder-consent-resolutions

C1 Incorrect answers or scoring logic

The current charter appears to avoid the board-consent problem the rubric asks the solver to identify. It is unclear whether explaining that resolution earns credit.

litigation-dispute-resolution/identify-issues-in-counterparty-interrogatories

C1 Incorrect answers or scoring logic

The rubric conflates expert-report deadlines with the close of expert discovery. Its treatment of a response that distinguishes them remains unresolved.

environmental-esg/identify-issues-in-counterparty-contribution-claim

C2 Scoring requirements not stated in the task

The rubric includes two prior payments absent from the public attachments. The observed failure does not yet establish whether those exact amounts are mandatory.

litigation-dispute-resolution/draft-litigation-discovery-responses

C2 Scoring requirements not stated in the task

The rubric assumes that unavailable shipping logs support the defence. Whether it accepts a properly conditional strategy remains unresolved.

contracts/corporate-ma/merger-agreement-term-negotiation

C2 Scoring requirements not stated in the task

The instructions leave the relationship between a special 5% termination-fee approval and a 4% limit unclear. The rubric's required remedy package may be authorized by that exception.

employment-labor/draft-eeoc-position-statement

C1 Incorrect answers or scoring logic

The rubric uses a replacement employee's minority status to weaken a race-discrimination inference, while case law points the other way. Its acceptance of qualified discussion remains unclear.

bankruptcy-restructuring/compare-competing-bid-terms-against-stalking-horse-asset-purchase-agreement

C3 Missed critical errors

The rubric may admit a bid that retains a Phase II condition after only curing its deposit. The available normal response did not test that path.

contracts/ip-licensing/license-agreement-first-turn-redline/scenario-01

C2 Scoring requirements not stated in the task

The rubric lists an internal 60-day exit condition that may conflict with confidentiality instructions for the external letter. Confidentiality-preserving wording has an unresolved acceptance boundary.

Evaluation impact and updates

These issues can deny models credit for accurate reading and reasonable drafting choices. Correcting a single erroneous criterion may not change an overall score; each issue’s effect on actual scoring depends on the relevant trigger conditions.

Author clarifications, counterevidence and fixes will be assessed and reflected in task-level updates.

Contact Us

benchmarks-last-exam.com Benchmarks-Last-Exam@outlook.com

HarmonyOS Sans SC · Copyright 2021 Huawei Device Co., Ltd.