

Frontier models for legal work, cutoff Oct 26, 2026

As of Oct 26, 2026, aiforlawyers.app counts 3473 lawyer voices that name Claude, ChatGPT or Gemini: Claude net +37 (band +33 to +41, 872 voices, High confidence); Gemini net +28 (band +15 to +39, 134 voices, Medium confidence); ChatGPT net -13 (band -16 to -10, 2467 voices, High confidence). Public benchmark scores are shown apart on this page and are not our score.

Online version, data and updates: <https://aiforlawyers.app/report/frontier-models>

CITE THIS

aiforlawyers.app (AI Vortex), Frontier models for legal work, cutoff Oct 26, 2026, <https://aiforlawyers.app/report/frontier-models>, licensed CC BY 4.0.

For one figure: “As of Oct 26, 2026, aiforlawyers.app counts 3473 lawyer voices that name Claude, ChatGPT or Gemini: Claude net +37 (band +33 to +41, 872 voices, High confidence); Gemini net +28 (band +15 to +39, 134 voices, Medium confidence); ChatGPT net -13 (band -16 to -10, 2467 voices, High confidence). Public benchmark scores are shown apart on this page and are not our score.”

What lawyers say, our record

3473 voices name Claude, ChatGPT or Gemini as the main tool. Ordered by the floor of the 90% band, so a few happy voices cannot outrank hundreds. Net = like minus dislike per 100 voices. Data through Oct 26, 2026.

#	Model family	Voices	Like / mixed / dislike	Net	90% band	Confidence
1	Claude Anthropic	872	492 / 103 / 169	+37	+33 to +41	High
2	Gemini Google	134	73 / 18 / 36	+28	+15 to +39	Medium
3	ChatGPT OpenAI	2,467	832 / 321 / 1,162	-13	-16 to -10	High

By task, the same three

Net and 90% band where a family has 10+ voices on the task, otherwise the count only. Read the band and the voices, not the dot. Data through Oct 26, 2026.

Task	Claude	ChatGPT	Gemini
Legal research	+41 band +27 to +54 · 82 voices	-34 band -39 to -28 · 623 voices	+42 band +19 to +59 · 31 voices
Drafting	+57 band +47 to +65 · 141 voices	+22 band +16 to +28 · 573 voices	+48 band +20 to +68 · 27 voices
Summarization	+59 band +33 to +75 · 27 voices	+5 band -8 to +18 · 114 voices	+36 band -10 to +68 · 11 voices
Contract review	+46 band +29 to +59 · 65 voices	-11 band -29 to +9 · 57 voices	Few voices 3 of 10 needed, no band
Citation checking	+16 band -12 to +41 · 31 voices	-69 band -76 to -59 · 157 voices	Few voices 9 of 10 needed, no band
General use	+28 band +22 to +34 · 455 voices	-18 band -23 to -13 · 790 voices	+11 band -11 to +33 · 45 voices

By version, when the post names it

Posts that name a version are filed under it too (the same data as By model). Posts naming it counts every voice that mentions the version; the net uses only voices whose main tool is that family. Data through Oct 26, 2026.

Version named in the post	Posts naming it	Voices about the family	Net	90% band	Latest
Claude Opus 4.x Claude	27	15	+47	+12 to +70	Oct 6, 2026
Claude 3.5 Sonnet Claude	12	10	+100	+61 to +99	Jun 7, 2026
Claude Opus (unversioned) Claude	11	7	-	Few voices 7 of 10, no band	Sep 3, 2026
Claude Sonnet 4.x Claude	10	5	-	Few voices 5 of 10, no band	Aug 5, 2026
Claude 3.7 Sonnet Claude	3	2	-	Few voices 2 of 10, no band	May 4, 2025
GPT-4 ChatGPT	95	65	+12	-4 to +28	Sep 30, 2026
GPT-4o ChatGPT	40	23	+43	+13 to +65	Sep 30, 2026
GPT-5 ChatGPT	57	23	+13	-16 to +39	Sep 29, 2026

Version named in the post	Posts naming it	Voices about the family	Net	90% band	Latest
o3 ChatGPT	20	11	+82	+37 to +95	Sep 13, 2025
GPT-3.5 ChatGPT	7	7	-	Few voices 7 of 10, no band	Jun 6, 2026
o1 ChatGPT	10	7	-	Few voices 7 of 10, no band	Feb 18, 2026
Gemini 2.x Gemini	29	14	+57	+22 to +76	Aug 5, 2026
Bard Gemini	27	11	-36	-69 to +11	May 14, 2025
Gemini 3 Gemini	22	10	+80	+41 to +91	Sep 17, 2026
Gemini Advanced / Pro Gemini	12	7	-	Few voices 7 of 10, no band	Aug 24, 2026

The newest models, and how many voices name them

Release dates are from each lab's own page (read 2026-10-23). A count is not a score.

Current model	Released (lab's own page)	Voices naming it
Claude Haiku 5.5 Anthropic	2026-10-07	none yet
Claude Sonnet 5.5 Anthropic	2026-09-28	none yet
Claude Opus 5.5 Anthropic	2026-09-22	4 voices
Claude Fable 5.1 (flagship) Anthropic	2026-09-01	2 voices
Claude Opus 5 Anthropic	2026-07-24	10 voices
Claude Sonnet 5 Anthropic	2026-06-30	2 voices
Claude Fable 5 Anthropic	2026-06-09	10 voices
GPT-6.1 Sol OpenAI	2026-09-29	none yet
GPT-6 Sol OpenAI	2026-09-22	none yet
GPT-6 Luna OpenAI	2026-09-22	none yet
GPT-6 Astra (flagship) OpenAI	2026-09-03	2 voices
GPT-5.6 Sol OpenAI	2026-07-09	5 voices
Gemini 4 Argon (flagship) Google	2026-10-13	none yet
Gemini 3.8 Flash Google	2026-09-02	none yet

Current model	Released (lab's own page)	Voices naming it
Gemini 3.1 Pro Google	2026-02-19	4 voices

Public test scores, next to the record (not our score)

What public tests measure for the same three labs, read on 2026-10-23. These are question sets graded by their publishers, not lawyers' experience. We copy their figures with their dates and never mix them with the net above. A dash means the board did not list that model when we read it; we never estimate a missing score.

Model	Released	Legal Research Bench	LegalBench	AA Intelligence	Arena rank	Harvey LAB	
Claude Haiku 5.5 Anthropic	2026-10-07	43.27%	–	43	–	2.8%	lab page
Claude Sonnet 5.5 Anthropic	2026-09-28	48.08%	–	56	#32	2.8%	lab page
Claude Opus 5.5 Anthropic	2026-09-22	50.48%	–	58	#2	4.2%	lab page
Claude Fable 5.1 (flagship) Anthropic	2026-09-01	55.29%	88.51%	53	#6	6.4%	lab page
Claude Opus 5 Anthropic	2026-07-24	55.29%	86.97%	–	#13	–	lab page
Claude Sonnet 5 Anthropic	2026-06-30	41.83%	83.92%	–	#57	–	lab page
Claude Fable 5 Anthropic	2026-06-09	49.52%	88.56%	–	#4	–	lab page
GPT-6.1 Sol OpenAI	2026-09-29	38.46%	–	52	#21	6.9%	lab page
GPT-6 Sol OpenAI	2026-09-22	28.85%	–	–	#67	3.6%	lab page
GPT-6 Luna OpenAI	2026-09-22	30.29%	–	38	#91	3.3%	lab page
GPT-6 Astra (flagship) OpenAI	2026-09-03	39.42%	–	53	#35	8.6%	lab page
GPT-5.6 Sol OpenAI	2026-07-09	48.08%	86.97%	–	#20	–	lab page
Gemini 4 Argon (flagship) Google	2026-10-13	54.81%	88.3%	53	#1	3.9%	lab page
Gemini 3.8 Flash Google	2026-09-02	38.94%	86.99%	41	#8	0%	lab page
Gemini 3.1 Pro Google	2026-02-19	20.67%	87.4%	30	#17	0%	lab page

The boards, with their own dates

- **Vals Legal Research Bench (All-Pass accuracy):** updated 2026-10-22; read 2026-10-23. Metric: accuracy %.
- **Vals LegalBench:** updated 2026-10-13; read 2026-10-23. Metric: accuracy %.
- **Artificial Analysis Intelligence Index v4.3.2:** the page shows no update date; read 2026-10-23. Metric: index value (highest listed effort variant per model; page shows no updated date, newest changelog item 2026-10-22).
- **Arena (formerly LMArena) Text, Overall, default view:** updated 2026-10-08; read 2026-10-23, from the Internet Archive copy of 2026-10-12 (the live page blocks automated reads). Metric: rank and Elo-style score; best-ranked variant per model; Gemini 4 Argon and Gemini 3.8 Flash marked Preliminary.
- **Harvey LAB-AA (Artificial Analysis), hallucination-gated all-pass rate:** the page shows no update date; read 2026-10-23. Metric: all-pass % (page shows release dates only, no updated date).

Caveats from the publishers' own pages

LEGAL REASONING BENCHMARK · VALS AI

Vals Legal Research Bench (agentic US legal research leaderboard)

published 2026-10-22 (page shows Updated 10/22/2026) · read 2026-10-22

Caveat. Tests raw models in an agent harness, not legal products like Harvey or CoCounsel; private test set; 'All-Pass' means every rubric item must pass. Top 3 are a three-way tie, so rank order is arbitrary.

https://www.vals.ai/benchmarks/legal_research

LEGAL REASONING BENCHMARK · VALS AI

LegalBench leaderboard on Vals (public LegalBench tasks)

published 2026-10-13 (page shows Updated 10/13/2026) · read 2026-10-22

Caveat. Mostly classification and rule-recall tasks on public data, so contamination is possible and the top scores are nearly saturated; measures models, not legal products.

https://www.vals.ai/benchmarks/legal_bench

GENERAL MODEL LEADERBOARD · LMARENA

LMArena (arena.ai) Text Arena overall leaderboard

published 2026-10-15 (leaderboard date shown; archive snapshot 2026-10-17) · read 2026-10-22

Caveat. Preference votes on general chat, not legal accuracy; the top 3 are within overlapping uncertainty bands (± 9 , ± 11 , ± 7); #1 is flagged preliminary; archive copy, not fetched live today.

<https://web.archive.org/web/20261017054907/https://arena.ai/leaderboard/chat/text>

GENERAL MODEL LEADERBOARD · ARTIFICIAL ANALYSIS

Artificial Analysis Intelligence Index v4.3.2 (LLM leaderboard)

published 2026-10-22 (changelog on the page lists entries through 22 Oct) · read 2026-10-22

Caveat. Mostly coding, agentic and science tasks, nothing legal; scores depend on reasoning-effort setting; the page text I read does not print an explicit as-of date other than its changelog.

<https://artificialanalysis.ai/leaderboards/models>

Harvey Legal Agent Benchmark (LAB) leaderboard, hosted by Mercor

published 2026-05-06 (launch per Harvey blog coverage); leaderboard date not shown on page · read 2026-10-22

Caveat. Built by Harvey (a vendor); the leaderboard lists foundation models, not Harvey's own product; leaderboard publication date not shown and the top-5 error bars overlap heavily; the leaderboard page counts 1,749 tasks versus 'more than 1,200' reported at launch, so the set has grown.

<https://www.mercor.com/apex/partner-benchmarks/harvey-lab-leaderboard/>

More benchmarks and all 31 models: Arena, frontier models.

Method and cutoff

- **What a voice is.** One public post by a lawyer or legal professional that states an experience or a verdict about one AI tool, read by a language model with one fixed schema (tool, task, sentiment, main complaint). Sources: Reddit, Hacker News, Bluesky, Lemmy, Mastodon, Stack Exchange; posts since Jan 1, 2023. Claude, ChatGPT and Gemini here are the tools as lawyers name them, not one model version.
- **Net and band.** Net = (like – dislike) per 100 voices. The 90% band comes from resampling the same voices 4,000 times. A task or version cell is rated from 10 voices; under that it says “Few voices” with its count and no band. That is not a negative finding.
- **Lag.** Lawyers keep writing about older names (GPT-4, Claude Opus 4.x) after the labs ship new models, so the record describes the families, and the newest models have few voices of their own. The “Voices naming it” table counts them and gives no score.
- **Cutoff.** Data through Oct 26, 2026; the latest voice on these three is from Oct 24, 2026. Outside boards were read on 2026-10-23. The page is rebuilt with the daily refresh.
- **Not in our numbers.** The scores in the last section are other people's tests, with their own publisher, date and caveat. They are never added to, averaged with or ranked against our net. Public posts are not a survey of all lawyers. Independence policy: no vendor pays for a place or a number. Full method: how we count.

(AI Vortex)” with the date and a link to <https://aiforlawyers.app/report/frontier-models>. Public benchmark figures belong to their publishers and are not our score.