FILING-EVIDENCE SECTORS · RESULTS PER MODEL

Llama-3.1-8B

10,000 questions over 200 filings of 100 companies in ten sectors, answered on one RTX 4090

Scorecard

64.4%
correct with a checkable quote, plain setup (5,799 of 9,002)

891
quietly wrong answers (confident-wrong) in 10,000, plain setup

89.6%
correct "not found" when the number was absent

35 Wh
GPU energy per 1,000 answers (₹0.28 at ₹8/kWh)

By sector

Plain setup (C1) and check + retry (C2), all 10,000 questions. Accuracy (Acc.) is the share of answerable questions answered correctly with a checkable quote; confident-wrong (CW) is the share of all questions answered wrongly with a quote that passes every check.

Sector	Acc. C1	95% CI	CW C1	Acc. C2	CW C2
Banks	65.8%	62.6–68.8	8.1%	69.2%	9.8%
Energy (power and utilities)	71.8%	68.7–74.6	7.8%	75.8%	8.9%
Healthcare	62.6%	59.3–65.7	9.6%	68.3%	11.0%
Insurance	62.8%	59.6–65.9	10.1%	66.4%	12.5%
Media	66.7%	63.5–69.7	7.5%	71.1%	9.8%
Metals and mining	63.1%	59.9–66.2	10.8%	67.2%	12.3%
Oil and gas	63.0%	59.8–66.1	15.5%	66.3%	17.1%
Real estate	64.8%	61.6–67.8	6.6%	69.2%	7.6%
Technology	64.3%	61.1–67.4	6.6%	69.7%	8.2%
Telecom	59.4%	56.2–62.6	6.5%	63.8%	7.9%
All sectors	**64.4%**	**63.4–65.4**	**8.9%**	**68.7%**	**10.5%**

Setups and checks

Each comparison is paired (the same questions) with an exact McNemar test; p is Holm-corrected within the model or the check. CW = confident-wrong. Thinking and the robustness checks use the 1,000-question subset.

Comparison	Correct	p	CW	p
Plain → check + retry (10,000)	5,799 → 6,185	< 0.001	891 → 1,051	< 0.001
Plain → Markdown tables, strict quote check (R1)	562 → 378	< 0.001	92 → 53	< 0.001
Plain → Markdown tables, pipes ignored (exploratory)	562 → 521	0.006	92 → 73	0.09
Plain → alternative prompt (R2)	562 → 524	0.05	92 → 85	1.00

Speed

Plain setup: median 3.8 s per answer (p95 6.4 s) at the runner's fixed concurrency, 14,412 answers per hour, 48 output tokens per answer on average.

← All models · Every answer of this model, with its raw reply, is in the dataset (https://huggingface.co/datasets/pramana-research/filing-evidence-sectors) under `results/llama-3.1-8b/`.

© 2026 Pramana Research · Text, data and figures: CC BY 4.0 · Code: Apache 2.0 · Legal

Model outputs are not investment advice. Check every figure against its source.

Pramana Research is an independent research group, not affiliated with any other company of a similar name.