FILING-EVIDENCE SECTORS · RESULTS PER MODEL

Gemma-3-12B

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

Scorecard

63.9%

correct with a checkable quote, plain setup (5,755 of 9,002)

1,043

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

82.1%

correct "not found" when the number was absent

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	62.0%	58.8–65.1	11.5%	64.6%	12.6%
Energy (power and utilities)	69.9%	66.8–72.8	9.3%	72.1%	10.3%
Healthcare	62.6%	59.3–65.7	13.4%	65.2%	14.8%
Insurance	64.7%	61.5–67.7	9.4%	67.8%	11.1%
Media	66.4%	63.3–69.5	8.5%	69.0%	9.5%
Metals and mining	58.4%	55.2–61.6	11.0%	61.8%	12.6%
Oil and gas	62.2%	59.0–65.3	16.5%	64.4%	18.2%
Real estate	58.9%	55.6–62.1	8.9%	62.8%	10.4%
Technology	65.8%	62.6–68.8	7.2%	69.2%	8.7%
Telecom	68.4%	65.3–71.4	8.6%	70.3%	9.9%
All sectors	**63.9%**	**62.9–64.9**	**10.4%**	**66.7%**	**11.8%**

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)	5,755 → 6,006	< 0.001	1,043 → 1,181	< 0.001
Plain → Markdown tables, strict quote check (R1)	558 → 437	< 0.001	95 → 76	0.08
Plain → Markdown tables, pipes ignored (exploratory)	558 → 517	0.009	95 → 90	1.00
Plain → alternative prompt (R2)	558 → 581	0.30	95 → 89	1.00

Speed

Plain setup: median 6.1 s per answer (p95 8.8 s) at the runner's fixed concurrency, 9,295 answers per hour, 60 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/gemma-3-12b/`.

© 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.