FILING-EVIDENCE SECTORS · RESULTS PER MODEL · SUPPLEMENTARY MODEL

Qwen3-8B

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

Scorecard

70.8%

correct with a checkable quote, plain setup (6,372 of 9,002)

942

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

87.7%

correct "not found" when the number was absent

43 Wh

GPU energy per 1,000 answers (₹0.34 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	68.8%	65.7–71.7	9.3%	69.9%	10.3%
Energy (power and utilities)	77.0%	74.1–79.6	6.7%	79.2%	7.3%
Healthcare	68.3%	65.2–71.3	11.0%	70.9%	11.4%
Insurance	70.8%	67.7–73.7	10.3%	72.3%	11.1%
Media	69.7%	66.6–72.6	9.3%	72.7%	9.6%
Metals and mining	69.8%	66.7–72.7	11.2%	71.3%	11.9%
Oil and gas	65.3%	62.1–68.3	17.0%	67.4%	17.6%
Real estate	76.2%	73.3–78.9	5.4%	77.4%	5.5%
Technology	71.1%	68.1–74.0	6.0%	72.7%	6.7%
Telecom	70.9%	67.8–73.8	8.0%	72.8%	8.6%
All sectors	**70.8%**	**69.8–71.7**	**9.4%**	**72.7%**	**10.0%**

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 → Markdown tables, strict quote check (R1)	605 → 551	< 0.001	103 → 87	0.08
Plain → Markdown tables, pipes ignored (exploratory)	605 → 599	1.00	103 → 97	1.00
Plain → alternative prompt (R2)	605 → 612	1.00	103 → 83	0.13

Speed

Plain setup: median 4.8 s per answer (p95 7.9 s) at the runner's fixed concurrency, 11,396 answers per hour, 57 output tokens per answer on average. Thinking: median 37.9 s per answer, 1095 output tokens 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/qwen3-8b/`.

© 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.