FILING-EVIDENCE SECTORS · RESULTS PER MODEL

Nemotron-Nano-9B-v2

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

Scorecard

45.9%
correct with a checkable quote, plain setup (4,128 of 9,002)

1,042
quietly wrong answers (confident-wrong) in 10,000, plain setup

99.4%
correct "not found" when the number was absent

51 Wh
GPU energy per 1,000 answers (₹0.41 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	47.7%	44.4–50.9	11.0%	47.7%	11.0%
Energy (power and utilities)	52.9%	49.6–56.1	8.9%	52.9%	8.9%
Healthcare	46.6%	43.3–49.8	10.0%	46.7%	10.2%
Insurance	41.9%	38.7–45.1	14.1%	41.9%	14.1%
Media	45.7%	42.4–48.9	10.4%	45.7%	10.4%
Metals and mining	47.8%	44.5–51.0	10.9%	47.9%	10.9%
Oil and gas	43.7%	40.5–46.9	14.3%	44.0%	14.4%
Real estate	43.7%	40.5–46.9	9.0%	43.9%	9.1%
Technology	45.1%	41.9–48.4	7.9%	45.4%	7.9%
Telecom	43.7%	40.5–46.9	7.7%	43.9%	7.8%
All sectors	**45.9%**	**44.8–46.9**	**10.4%**	**46.0%**	**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)	4,128 → 4,140	0.001	1,042 → 1,047	0.06
Plain → thinking (subset)	412 → 519	< 0.001	92 → 68	0.06
Plain → Markdown tables, strict quote check (R1)	412 → 255	< 0.001	92 → 45	< 0.001
Plain → Markdown tables, pipes ignored (exploratory)	412 → 406	1.00	92 → 79	0.64
Plain → alternative prompt (R2)	412 → 339	< 0.001	92 → 145	< 0.001

Speed

Plain setup: median 6.1 s per answer (p95 10.3 s) at the runner's fixed concurrency, 9,348 answers per hour, 53 output tokens per answer on average. Thinking: median 16.1 s per answer, 626 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/nemotron-nano-9b-v2/`.

© 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.