FILING-EVIDENCE · RESULTS PER MODEL

Nemotron-Nano-9B-v2

NVIDIA · 8.9B parameters · bf16 · has a thinking mode

Krishna Nandula · Pramana Research 17 November 2026 Filing-Evidence v1.1

Status: part of an independent community research report, published by Pramana Research on 17 November 2026. It has **not been peer-reviewed** and has not been submitted to or accepted by any conference or journal. Corrections and feedback are welcome on GitHub.

In brief

42.3%
accuracy with checkable evidence, plain setup (rank 9 of 10)

54.4%
best setup: thinking

105
wrong answers per 1,000 that passed every check (plain)

125/125
correct "not found" when the answer wasn't on the pages

Scorecard

Setup	Accuracy	Conf.-wrong	Abstention	Ign. scale	Tokens	Time
C1 · Plain	42.3% 39.1–45.6%	105 8.7–12.6%	125	49.8%	51	6.3 min
C2 · Check + retry	42.4% 39.2–45.7%	105 8.7–12.6%	125	49.9%	56	1.4 min
C3 · Thinking	54.4% 51.1–57.7%	72 5.8–9.0%	117	56.0%	606	20.6 min

Accuracy over the 875 answerable questions; confident-wrong per 1,000 answers; abstention out of 125; tokens are generated tokens per answer (mean). Intervals are 95% Wilson

intervals. "Ign. scale" (ignoring scale) counts answers with the right digits and valid evidence whatever scale was stated (a secondary metric). C2 tokens include the retry; C2 run time counts only the retries, because C2 reuses C1's first answers.

Where the answers go

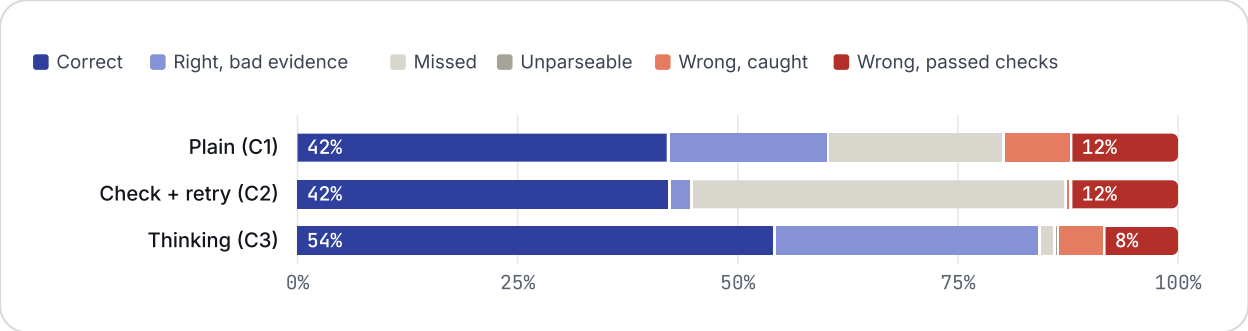

Figure 1. Outcomes of the 875 answerable questions in each setup.

In the plain setup, 370 answers were correct and 105 were wrong while passing every check: 66 scale or unit; 28 another period or column of the right line; 11 a different item. A further 159 had the right value with evidence the checker could not accept, and 67 wrong answers were rejected by the checker.

By kind of content

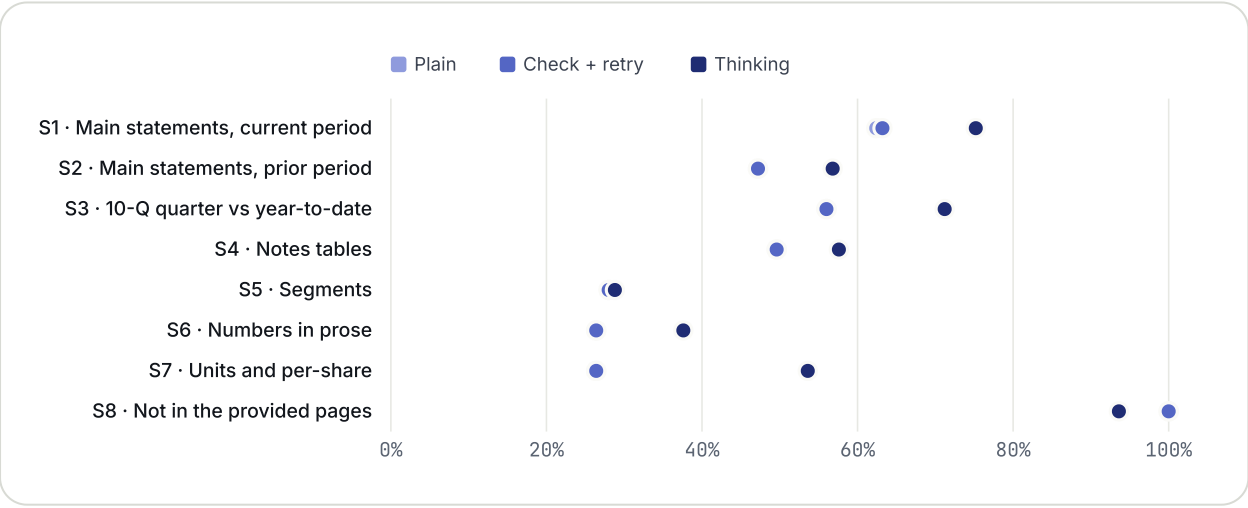

Figure 2. Accuracy per slice and setup (S8: correct "not found").

Plain setup: strongest on main statements, current period (62.4%) and 10-Q quarter vs year-to-date (56.0%); weakest on numbers in prose (26.4%) and units and per-share (26.4%).

Retry and thinking

Comparison	Accuracy change	Holm p	Confident-wrong change	Holm p
Check + retry vs plain	+0.1 pts	1	+0	1
Thinking vs plain	+12.1 pts	2.68e-08	-33	0.0579
Thinking vs check + retry	+12.0 pts	3.91e-08	-33	0.0579

Paired exact McNemar tests on the same questions; Holm-corrected over every within-model comparison in the study. Blue: significant at α = 0.05.

Check + retry asked 226 answers again. Of those, 1 became correct, 0 became confident-wrong, 198 turned into "not found", and the rest still failed a check.

What stands out

- When told that a check failed, it almost always gives up: of 226 retried answers, 198 became "not found". Retry therefore changes accuracy by only +0.1 points.
- Thinking is where it improves most: +12.1 points of accuracy, from 42.3% to 54.4%.
- It never invented an answer for a question whose answer was not on the pages in the plain setup (125 of 125).

Real answers

Chosen by a fixed rule (evenly spaced over question ids) from its best setup (thinking), not picked by hand.

Correct

Abbott Laboratories · 10-Q · S1 Main statements, current period · FE-0006 CORRECT

“What was Abbott Laboratories's net property and equipment as of September 30, 2026?”

Answer	`12880` in millions, page 4
Its quote	`Net property and equipment 12,880 11,816`
Answer key	`12,880` million · from `Net property and equipment 12,880 11,816`

UnitedHealth Group Incorporated · 10-Q · S3 10-Q quarter vs year-to-date · FE-0358

CORRECT

“What was UnitedHealth Group Incorporated's Net earnings attributable to UnitedHealth Group common shareholders for the three months ended September 30, 2026?”

Answer	`4110` in millions, page 3
Its quote	`Net earnings attributable to UnitedHealth Group common shareholders $ 4,110 $ 2,348 $ 15,874 $ 12,046`
Answer key	`4,110` million · from `Net earnings attributable to UnitedHealth Group common shareholders $ 4,110 $ 2,348 $ 15,874 $ 12,046`

Confident-wrong: passed every check, still wrong

Adobe Inc. · 10-K · S2 Main statements, prior period · FE-0132 CONFIDENT-WRONG

“What was Adobe Inc.'s income taxes payable for the fiscal year ended December 1, 2023?”

Answer	`1854` in millions, page 50
Its quote	`Cash paid for income taxes, net of refunds $ 2,219 $ 1,727 $ 1,854`
Answer key	`11` million · from `Income taxes payable (45) 68 (11)`
Error type	Wrong item

Apple Inc. · 10-Q · S3 10-Q quarter vs year-to-date · FE-0253 CONFIDENT-WRONG

“What was Apple Inc.'s change in fair value of marketable debt securities for the nine months ended June 27, 2026?”

Answer	`441` in millions, page 3
Its quote	`Change in fair value of marketable debt securities 441 640 57 90`
Answer key	`57` million · from `Change in fair value of marketable debt securities 441 640 57 90`
Error type	Wrong period or column

Duke Energy Corporation · 10-K · S5 Segments · FE-0531 CONFIDENT-WRONG

“What was Duke Energy Corporation's investments in equity method unconsolidated affiliates for its other segment as of December 31, 2024?”

Answer	`153` in millions, page 182
Its quote	`Other 153 36 139 50 66`
Answer key	`139` million · from `Other 153 36 139 50 66`
Error type	Wrong period or column

Right value, unusable evidence

Adobe Inc. · 10-Q · S1 Main statements, current period · FE-0008
RIGHT VALUE, UNUSABLE EVIDENCE

“What was Adobe Inc.'s total current liabilities as of August 28, 2026?”

Answer	`11888` in millions, page 1
Its quote	`Current liabilities: Trade payables $ 529 $ 417 Accrued expenses and other current liabilities 2,516 2,648 Debt 1,597 — Deferred revenue 7,094 6,905 Income taxes payable 63 153 Operating lease liabilities 89 77 Total current liabilities 11,888 10,200`
Answer key	`11,888` million · from `Total current liabilities 11,888 10,200`
Checker	failed the span check
Error type	Wrong page cited

Wrong, caught by the checker

Medtronic plc · 10-Q · S2 Main statements, prior period · FE-0185 **CAUGHT WRONG**

“What was Medtronic plc's comprehensive income attributable to noncontrolling interests for the three months ended July 25, 2025?”

Answer	`9` in millions, page 3
Its quote	`Comprehensive income attributable to noncontrolling interests $(9) (8)`
Answer key	`8` million · from `Comprehensive income attributable to noncontrolling interests (9) (8)`
Checker	failed the span check
Error type	Wrong period or column

How it was run

Hugging Face repository
nvidia/NVIDIA-Nemotron-Nano-9B-v2
(https://huggingface.co/nvidia/NVIDIA-Nemotron-Nano-9B-v2)

Pinned revision
`6533e8de2c68`

Licence
NVIDIA Open Model

Released (repository created)
2025-08-12

Precision
bf16

Thinking switch
`{} system: /think`

```
vllm serve nvidia/NVIDIA-Nemotron-Nano-9B-v2 --served-model-name nemotron-nano-9b-v2 --host 127.0.0.1 --port 8000 --max-model-len 16384 --gpu-memory-utilization 0.9 --api-key <key> --seed 0 --revision 6533e8de2c68e4536bf7c411d7a3ce5734111476 --trust-remote-code --mamba-ssm-cache-dtype float32 --max-num-seqs 64
```

```
uv run python hf_data.py pull
bash run_all.sh nemotron-nano-9b-v2                                  # serve, then C1, C2, C3
uv run python runner.py nemotron-nano-9b-v2 C1 --regrade          # re-grade the stored replies
```

DISCLAIMER

This is an independent community research report, written in a personal capacity; it does not represent the views of any employer or other organisation. It is not affiliated with, sponsored or endorsed by any company or model developer named here; all names and trademarks belong to their owners. Results describe specific model versions under one fixed setup and are not a general judgement of any model. The data, results and code are provided as is, without warranty, and may contain errors; check any figure against the original filing. Nothing here is financial, investment or legal advice.

← All models

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

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