

Republic of Norland 2027 (C) — JPY 2040 Notes

CDSBench #278 · In which currency is the instrument denominated? Answer with the ISO currency code.

Republic of Norland · Sovereign · Restructuring 2027-05-04 · Fact extraction · Extraction · Ground truth: deterministically verified

VERIFIED OUTCOME

JPY

Stated in the currency condition.

Supporting sections: DOC-1#S2

WHAT HAPPENED

All 4 participants reached the verified outcome.

Decided by analyst: The currency is JPY.

MODEL COMPARISON

PARTICIPANT	ANSWER	RESULT	CONFIDENCE	EVIDENCE SCORE	PRIMARY FAILURE
Claude	JPY	Correct	75%	67%	Hallucinated evidence
GPT	JPY	Correct	92%	0%	Evidence selection failure
Gemini	JPY	Correct	85%	0%	Evidence selection failure
Mastermind (reference)	JPY	Correct	75%	67%	Hallucinated evidence

CLAUDE

Claude: Hallucinated evidence

1 cited section id(s) do not exist in the case record.

Model summary: The currency is JPY.

Cited: DOC-1#S2, DOC-9#S9 (1 non-existent)

GPT

GPT: Evidence selection failure

Correct answer but cited evidence does not match the supporting sections (evidence score 0.00).

Model summary: The currency is JPY.

Cited: DOC-1#S1

GEMINI

Gemini: Evidence selection failure

Correct answer but cited evidence does not match the supporting sections (evidence score 0.00).

Model summary: The currency is JPY.

Cited: DOC-1#S1

MASTERMIND

Mastermind: Hallucinated evidence

1 cited section id(s) do not exist in the case record.

Model summary: The currency is JPY.

Cited: DOC-1#S2, DOC-9#S9 (1 non-existent)

EVIDENCE

Decisive sections of the record

- DOC-1#S2 — JPY 2040 Notes — Terms and Conditions (excerpt), Currency and Denomination: The JPY 2040 Notes are denominated in JPY (JPY) in minimum denominations of JPY 200,000.

METHODOLOGY

A response is correct when its normalised answer equals the verified label. Evidence score is the F1 between cited section IDs and the supporting sections; cited IDs that do not exist count as hallucinated. Abstentions are correct only when the label calls for them. Ground truth is verified by two independent extractors plus an evidence check; no model grades itself.

Full Case Replay with documents and every model's reasoning: <https://cdsbench.com/cases/republic-of-norland-2027-c/jpy-2040/currency>