

United Provinces of Kestrel 2027 — JPY 2040 Notes

CDSBench #80 · What is the maturity date of the instrument? Answer as YYYY-MM-DD.

United Provinces of Kestrel · Sovereign · Restructuring 2027-09-25 · Fact extraction · Extraction · Ground truth: deterministically verified

VERIFIED OUTCOME

2040-09-25

Stated in the maturity condition.

Supporting sections: DOC-1#S3

WHAT HAPPENED

All 4 participants reached the verified outcome.

Decided by analyst: The maturityDate is 2040-09-25.

MODEL COMPARISON

PARTICIPANT	ANSWER	RESULT	CONFIDENCE	EVIDENCE SCORE	PRIMARY FAILURE
Claude	2040-09-25	Correct	75%	100%	—
GPT	2040-09-25	Correct	92%	0%	Evidence selection failure
Gemini	2040-09-25	Correct	85%	0%	Evidence selection failure
Mastermind (reference)	2040-09-25	Correct	75%	100%	—

GPT

GPT: Evidence selection failure

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

Model summary: The maturityDate is 2040-09-25.

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 maturityDate is 2040-09-25.

Cited: DOC-1#S1

EVIDENCE

Decisive sections of the record

- DOC-1#S3 — JPY 2040 Notes — Terms and Conditions (excerpt), Maturity: The JPY 2040 Notes mature on 2040-09-25. Interest is payable semi-annually in arrear.

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/united-provinces-of-kestrel-2027/jpy-2040/maturity-date>