

United Provinces of Kestrel 2026 (B) — GBP 2033 Notes

CDSBench #220 · Does the instrument's maturity satisfy a maximum maturity of 30 years measured from the credit event date (2026-11-13)?

United Provinces of Kestrel · Sovereign · Restructuring 2026-11-13 · Rule application · Maturity · Ground truth: deterministically verified

VERIFIED OUTCOME

Yes

Maturity 2033-11-13 vs limit 2056-11-13.

Supporting sections: DOC-1#S3, DOC-2#S1

WHAT HAPPENED

All 4 participants reached the verified outcome.

Decided by analyst: Based on the maturityDate provision, the answer is YES.

MODEL COMPARISON

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

GPT

GPT: Evidence selection failure

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

Model summary: Based on the maturityDate provision, the answer is YES.

Cited: DOC-1#S1

GEMINI

Gemini: Hallucinated evidence

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

Model summary: Based on the maturityDate provision, the answer is YES.

Cited: DOC-1#S1, DOC-9#S9 (1 non-existent)

EVIDENCE

Decisive sections of the record

- DOC-1#S3 — GBP 2033 Notes — Terms and Conditions (excerpt), Maturity: The GBP 2033 Notes mature on 2033-11-13. Interest is payable semi-annually in arrear.
- DOC-2#S1 — Determinations Committee notice — United Provinces of Kestrel, Determination: The Determinations Committee resolved that a Restructuring Credit Event occurred with respect to United Provinces of Kestrel on 2026-11-13.

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-2026-b/gbp-2033/maturity-limit>