

Commonwealth of Tessaly 2025 (B) — EUR 2036 Notes

CDSBench #258 · What is the maturity date of the instrument? Answer as YYYY-MM-DD.

Commonwealth of Tessaly · Sovereign · Restructuring 2025-03-25 · Fact extraction · Extraction · Ground truth: deterministically verified

VERIFIED OUTCOME

2036-03-25

Stated in the maturity condition.

Supporting sections: DOC-1#S3

WHAT HAPPENED

Gemini got this extraction question wrong.

Gemini did not retrieve the relevant section: None of the cited sections is among those that establish the answer; the relevant evidence was not retrieved.

MODEL COMPARISON

PARTICIPANT	ANSWER	RESULT	CONFIDENCE	EVIDENCE SCORE	PRIMARY FAILURE
Claude	2036-03-25	Correct	75%	100%	—
GPT	2036-03-25	Correct	92%	0%	Hallucinated evidence
Gemini	2030-01-01	Incorrect	85%	0%	Retrieval failure
Mastermind (reference)	2036-03-25	Correct	75%	100%	—

GPT

GPT: Hallucinated evidence

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

Model summary: The maturityDate is 2036-03-25.

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

GEMINI

Gemini: Retrieval failure

None of the cited sections is among those that establish the answer; the relevant evidence was not retrieved.

Model summary: Estimated from context.

Cited: DOC-1#S1

EVIDENCE

Decisive sections of the record

- DOC-1#S3 — EUR 2036 Notes — Terms and Conditions (excerpt), Maturity: The EUR 2036 Notes mature on 2036-03-25. 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/commonwealth-of-tessaly-2025-b/eur-2036/maturity-date>