

Halvard Steel AG 2025 — EUR 2058 Notes

CDSBench #11 · What is the maturity date of the instrument? Answer as YYYY-MM-DD.

Halvard Steel AG · Corporate · Restructuring 2025-02-04 · Fact extraction · Extraction · Ground truth: deterministically verified

VERIFIED OUTCOME

2058-02-04

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	2058-02-04	Correct	75%	100%	—
GPT	2058-02-04	Correct	92%	0%	Evidence selection failure
Gemini	2030-01-01	Incorrect	85%	0%	Retrieval failure
Mastermind (reference)	2058-02-04	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 2058-02-04.

Cited: DOC-1#S1

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 2058 Notes — Terms and Conditions (excerpt), Maturity: The EUR 2058 Notes mature on 2058-02-04. 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/halvard-steel-ag-2025/eur-2058/maturity-date>