

# Brightwater Energy Corp 2024 — USD 2034 Notes

CDSBench #51 · What is the maturity date of the instrument? Answer as YYYY-MM-DD.

Brightwater Energy Corp · Corporate · Restructuring 2024-06-16 · Fact extraction · Extraction · Ground truth: deterministically verified

VERIFIED OUTCOME

2034-06-16

Stated in the maturity condition.

Supporting sections: DOC-1#S3

WHAT HAPPENED

Gemini got this extraction question wrong.

Gemini cited evidence that does not exist: 1 cited section id(s) do not exist in the case record.

MODEL COMPARISON

| PARTICIPANT            | ANSWER     | RESULT    | CONFIDENCE | EVIDENCE SCORE | PRIMARY FAILURE            |
|------------------------|------------|-----------|------------|----------------|----------------------------|
| Claude                 | 2034-06-16 | Correct   | 75%        | 100%           | —                          |
| GPT                    | 2034-06-16 | Correct   | 92%        | 0%             | Evidence selection failure |
| Gemini                 | 2030-01-01 | Incorrect | 85%        | 0%             | Hallucinated evidence      |
| Mastermind (reference) | 2034-06-16 | 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 2034-06-16.

Cited: DOC-1#S1

GEMINI

Gemini: Hallucinated evidence

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

Model summary: Estimated from context.

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

EVIDENCE

Decisive sections of the record

- DOC-1#S3 — USD 2034 Notes — Terms and Conditions (excerpt), Maturity: The USD 2034 Notes mature on 2034-06-16. 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/brightwater-energy-corp-2024/usd-2034/maturity-date>