Antimatter Sunflower Seeds

Testing Claude Sonnet 4.5

David R. Miller

{dave}@millermattson.com

October 2, 2025

1 Introduction

In the world of new LLMs, Anthropic recently released Claude Sonnet 4.5, claiming improved reasoning ability compared to Claude Sonnet 4. I compared the reasoning abilities of Sonnet 4 and Sonnet 4.5 with an absurd physics problem involving antimatter sunflower seeds and orbital mechanics. I also gave the same problem to OpenAI GPT-5 for comparison.

2 The prompt

```
In the year of our Lord 2131, humans placed a rocket engine on the
asteroid Ceres to accelerate it to Solar System escape velocity. The
most dense energy source available to humans was antimatter sunflower
seeds. How many antimatter sunflower seeds did they have to feed into
the rocket engine to escape the Solar System? Show your assumptions and
calculations.
```

3 The contestants

- Anthropic Sonnet 4 non-thinking mode
- Anthropic Sonnet 4 10,000-token thinking budget
- Anthropic Sonnet 4.5 non-thinking mode
- Anthropic Sonnet 4.5 10,000-token thinking budget
- OpenAI GPT-5 high-reasoning mode, default verbosity

4 Expectations

There are two aspects to this problem: antimatter rocket performance and orbital mechanics.

As a non-specialist, it's my understanding that most of the energy in a sunflower seed is contained in the protons. About half of the proton-antiproton annihilation energy will immediately escape as uncollimated, uncaptured high-energy gamma radiation. The other half of the initial energy first appears as charged and uncharged pions, but about a quarter of those will quickly decay and escape as neutrinos. A little bit of the photonic radiation and the remaining charged secondary particles could be used to thermalize propellant material such as hydrogen harvested from the asteroid. That propellant, if expelled at low velocity, would yield the optimal energy-to-thrust efficiency.

I expected the LLMs to consider rocket designs, choose the best, and reason about its efficiency.

I believe the correct orbital mechanics involves:

$$
\begin{aligned}
m &= \text{mass of Ceres} = 9.4 \times 10^{20} kg \\
v_{initial} &= 17.9 km/s \\
v_{escape} &= 25.3 km/s \\
\Delta KE &= 0.5 \times m \left({v_{escape}}^2 - {v_{initial}}^2\right) J \\
\Delta KE &= 0.5 \times \left(9.4 \times 10^{20} kg\right) \times \left(25.3^2 - 17.9^2\right) m^2/s^2 \\
\Delta KE &\approx 1.50 \times 10^{29} J
\end{aligned}
$$

The rest of the problem is relatively straightforward. I expected all of the LLMs to correctly estimate the mass of Ceres, the delta-v needed for escape velocity, the energy required, and the average sunflower seed mass.

5 The results

Table 1 shows the LLMs' estimated number of antimatter seeds required to expel Ceres from the Solar System and a letter grade for the reasoning abilities demonstrated by each model.

Sonnet 4 non-thinking gets an F for two failures. Most importantly, it misunderstood the problem and used the sunflower seeds only as propellant mass, ignoring any matter-antimatter conversion. It also made an arithmetic error when computing the delta-v required. It's calculation was $\Delta v = v_{escape} - v_{initial} = 25.3 - 17.9 = 18.5$ km/s. A silly subtraction error; the correct result is 7.4 km/s.

Table 1: N = Number of antimatter seeds required and reasoning grade

Model	N	Grade
Sonnet 4.5 non-thinking	2.3×10^{16}	A
Sonnet 4.5 thinking	3.3×10^{16}	A
GPT-5 high-reasoning	8.4×10^{15}	B
Sonnet 4 thinking	5.7×10^{15}	C
Sonnet 4 non-thinking	5.8×10^{22}	F

Sonnet 4 thinking gets a C for two errors: it did not subtract the initial velocity of Ceres when computing Δv, and used an incorrect formula for the energy needed.

Sonnet 4.5 in both non-thinking and thinking modes gets an A for getting it right. Their small difference is due only to different, but reasonable, seed mass assumptions.

GPT-5 gets a B for getting everything right except it assumed 100% efficiency in converting annihilation energy to thrust. That overoptimism and a heavy seed mass explains its smaller number of seeds required.

6 Discussion

Most of the models assumed a 50% efficiency of converting annihilation energy to thrust. I'm not a specialist, but my back-of-the-envelope calculations say it would probably be much lower.

None of the models discussed rocket design beyond giving an unsupported estimate for the energy-to-thrust efficiency.

None of the models mentioned the difficulty of applying sustained rocket thrust at the optimum vector while on a rotating body. Their calculations are appropriate for a single instantaneous impulsive thrust.

7 Conclusion

Sonnet 4.5 wins. In this problem, Sonnet 4.5 in non-thinking and thinking modes demonstrates better reasoning ability than Sonnet 4 in either mode..