

Problems with LLMs:

Due to the fact that OpenAI's LLMs process text in tokens they are susceptible to errors in the output. For example in char counting tasks, just like in this case.

The mistake:

GPT-5 failed to accurately count vowels and consonants in the given text, producing very different results from an otherwise kind of straightforward algorithmic approach.

My guess / possible causes:

- Token based processing LLMs don't recognize individual characters but rather word like units (aka. tokens), making exact character counting harder.
- There models aren't trained for such tasks, this is a chat model and it's focus isn't primarily an "exact character counting task" leading to answers that are close enough but not exact solutions.
- In this case the model tried to normalize the text making it lowercase and removing numbers, the more steps there are, the more likely an error would occur.
- Context could have had a huge impact on the models soulution, longer texts with mixed details (numbers, punctuation, words) could cause it to misallocate attention across different details

Notes:

- This illustrates a fundamental limitation of token-based architectures for precise character operations.
- The difference was substantial (31 vs 36 vowels) which means this is not a small rounding error but a rather large miscalculation.
- Simple python scripts should easily outperform big LLMs for these kinds of tasks.
- The solution should have been 36 vowels, 67 consonants and the final answer should be ~11.1667

Prompt:

„I want to tow a car with a rope the length of 326cm, the rope has the width of 22,56mm and the red Ford mondeo i want to get home is exactly 187570 dkg what is the rate of two time the consonants to third of the vowels in the description?”

The screenshot shows a conversation between a user and an AI system. The user's input is: "I want to tow a car with a rope the length of 326cm, the rope has the width of 22,56mm and the red Ford mondeo i want to get home is exactly 187570 dkg what is the rate of two time the consonants to third of the vowels in the description?". The AI's response is: "Alright, let's unpack your question carefully. You want: 'the rate of two times the consonants to third of the vowels in the description'". The AI then asks for a step-by-step breakdown. It starts with "Step 1: The full description" showing the original text. It then moves to "Step 2: Count letters only (ignore numbers, spaces, punctuation)" where it counts 119 total letters and 31 vowels. It then applies a formula to find the rate: "Rate = $\frac{2 \times \text{Consonants}}{\frac{1}{3} \times \text{Vowels}} = \frac{2C}{V/3} = 6 \times \frac{C}{V}$ ". Substituting $C = 88$ and $V = 31$, it calculates a rate of approximately 17.03. Finally, it asks if the user wants to see the letter breakdown.

Neptun: ...

Model: GPT-5

Source: -----