

ChatGPT struggles with Wordle puzzles, which says a lot about how it works

March 27 2023, by Michael G. Madden



Credit: AI-generated image ([disclaimer](#))

The AI chatbot known as ChatGPT, developed by the company OpenAI, has caught the public's attention and imagination. Some applications of the technology [are truly impressive](#), such as its ability to [summarize complex topics](#) or to [engage in long conversations](#).

It's no surprise that other AI companies [have been rushing](#) to release their own [large language models \(LLMs\)](#)—the name for the technology underlying chatbots like ChatGPT. Some of these LLMs will be incorporated into other products, such as search engines.

With its impressive capabilities in mind, I decided to test the chatbot on [Wordle](#)—the [word game](#) from the New York Times—which I have been playing for some time. Players have six goes at guessing a five-[letter](#) word. On each guess, the game indicates which letters, if any, are in the correct positions in the word.

Using the latest generation, [called ChatGPT-4](#), I discovered that its performance on these puzzles was surprisingly poor. You might expect word games to be a piece of cake for GPT-4. LLMs are "trained" on text, meaning they are exposed to information so that they can improve at what they do. ChatGPT-4 was trained on about 500 billion words: all of Wikipedia, all public-domain books, huge volumes of scientific articles, and text from many websites.

AI chatbots could play a major role in our lives. Understanding why ChatGPT-4 struggles with Wordle provides insights into how LLMs represent and work with words—along with the limitations this brings.

First, I tested ChatGPT-4 on a Wordle puzzle where I knew the correct locations of two letters in a word. The pattern was "#E#L#", where "#" represented the unknown letters. The answer was the word "mealy".

Five out of ChatGPT-4's six responses failed to match the pattern. The responses were: "beryl", "feral", "heral", "merle", "revel" and "pearl".

With other combinations, the chatbot sometimes found valid solutions. But, overall, it was very hit and miss. In the case of a word fitting the pattern "##OS#", it found five correct options. But when the pattern was

"#R#F#", it proposed two words without the letter F, and a word—"Traff"—that isn't in dictionaries.

Under the hood

At the core of ChatGPT is a [deep neural network](#): a complex mathematical function—or rule—that maps inputs to outputs. The inputs and outputs must be numbers. Since ChatGPT-4 works with words, these must be "translated" to numbers for the [neural network](#) to work with them.

The translation is performed by a [computer program](#) called a [tokenizer](#), which maintains a huge list of words and letter sequences, called "tokens". These tokens are identified by numbers. A word such as "friend" has a token ID of 6756, so a word such as "friendship" is broken down into the tokens "friend" and "ship". These are represented as the identifiers 6756 and 6729.

When the user enters a question, the words are translated into numbers before ChatGPT-4 even starts processing the request. The [deep neural network](#) does not have access to the words as text, so it cannot really reason about the letters.

Poem task

ChatGPT-4 is good at working with the first letters of words. I asked it to write a poem where the opening letter of each line spelled out "I love robots". Its response was surprisingly good. Here are the first four lines:

I am a fan of gears and steel

Loving their movements, so surreal,

Over circuits, they swiftly rule

Vying for knowledge, they're no fool,

[The training data for ChatGPT-4](#) includes huge numbers of textbooks, which often include alphabetical indices. This could have been enough for GPT-4 to have learned associations between words and their first letters.

The tokenizer also appears to have been modified to recognize requests like this, and seems to split a phrase such as "I Love Robots" into individual tokens when users enter their request. However, ChatGPT-4 was not able to handle requests to work with the last letters of words.

ChatGPT-4 is also bad at palindromes. Asked to produce a palindrome phrase about a robot, it proposed "a robot's sot, orba", which does not fit the definition of a palindrome and relies on obscure words.

However, LLMs are relatively good at generating other [computer programs](#). This is because their [training data](#) includes many websites devoted to programming. I asked ChatGPT-4 to write a program for working out the identities of missing letters in Wordle.

The initial program that ChatGPT-4 produced had a bug in it. It corrected this when I pointed it out. When I ran the program, it found 48 valid words matching the pattern "#E#L#", including "tells", "cells" and "hello". When I had previously asked GPT-4 directly to propose matches for this pattern, it had only found one.

Future fixes

It might seem surprising that a large language model like ChatGPT-4 would struggle to solve simple word puzzles or formulate palindromes,

since the training data includes almost every word available to it.

However, this is because all text inputs must be encoded as numbers and the process that does this doesn't capture the structure of letters within words. Because [neural networks](#) operate purely with numbers, the requirement to encode words as numbers will not change.

There are two ways that future LLMs can overcome this. First, ChatGPT-4 knows the first letter of every word, so its training data could be augmented to include mappings of every letter position within every word in its dictionary.

The second is a more exciting and general solution. Future LLMs could generate code to solve problems like this, as I have shown. A recent paper demonstrated [an idea called Toolformer](#), where an LLM uses external tools to carry out tasks where they normally struggle, such as arithmetic calculations.

We are in the early days of these technologies, and insights like this into current limitations can lead to even more impressive AI technologies.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: ChatGPT struggles with Wordle puzzles, which says a lot about how it works (2023, March 27) retrieved 24 April 2024 from <https://techxplore.com/news/2023-03-chatgpt-struggles-wordle-puzzles-lot.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.