Understanding Large Language Models

Blaine Swieder

18 July 2025

1 Introduction

Large Language Models (LLMs) are essentially very complicated prediction machines. LLMs are trained on vast amounts of text data - be it books, articles, websites, conversations - learning patterns, grammar, facts, and relationships between words.

2 How do LLMs Process Text?

When you type into a LLM such OpenAI's ChatGPT, Google's Gemini, or Anthropic's Claude, here is a breakdown of what happens.

- **Tokenization**: A prompt is broken into small units called *tokens* (words, parts of words, punctuation).
- **Pattern Recognition**: The LLM looks at your tokens and predicts the most probable next token based on its training.
- **Generation**: The LLM adds that predicted token, the predicts the next, building a response word by word until completion.

To summarize, think of the LLM as a genius auto-complete. It is not "intelligent" as mankind, but it is still predicting the most statistically probable sequence of words. This is why having a clear prompt and instructions are essential as they guide the prediction process.

3 The Key Capabilities of LLMs

The following is a list of some of the capabilities of Large Language Models (LLMs):

- **Understanding and Generating Human Language**: LLMs are able to write, to summarize, to translate, can answer questions, and generate creative text.

- **Pattern Recognition**: LLMs are execellent at identifying and continuing patterns (this is why ***few-shot learning*** works!).
- **Vast Knowledge Base**: Has a vast amount of training data.
- **Following Instructions**: It excels at adhering to formats, tones, and constraints when prompts are clear.
- **Reasoning (to a certain extent)**: Can perform logical deduction, common sense reasoning, and break down complex problems.

However, LLMs are not without their limitations.

4 The Limitations of LLMs

Below are some of the disadvantages of large language models:

- **Lack of True Understanding**: The LLM has no beliefs, emotions, and/or experiences. Responses are statistical, not genuine comprehension.
- **Hallucinations**: LLMs can also confidently generate factually incorrect info, so always fact-check critical information.
- **Outdated Knowledge**: There is knowledge cut-off at last training update and there is no access to the internet unless it is integrated.
- **Bias from Training Data**: The LLM may reflect biases present in their training data (especially political!), be sure to fact-check.
- **Sensitive to Prompt Wording**: Small phrasing changes can definitely effect how the LLM output is generated.
- **No Memory (by default)**: Each prompt is a new interaction, and the "memory" needs to be explicitly provided in the prompt for the AI agent.

By understanding all these points, you will gain an understanding of how to properly conduct your prompt engineering with realistic expectations. As a result, you will design prompts that play into the LLM's strengths rather than weaknesses.

All of this is from an Instagram post I saw, and I transcribed in more formal language.