Best Prompt Techniques for Best LLM Responses

Jules S. Damji Follow Published in The Modern Scientist · 7 min read · Nov 19, 2023

Better prompts is all you need for better responses

Introduction

The notion of a prompt is not new. This idea of offering a prompt permeates the field of natural language processing (NLP) and human-computer interaction (HCI) as a means to an experiment, a minimal investigation to offer a bite-sized (tiny) computer programming to take an initial step, given a specific context. All these are distribution prompts as requests to elicit a desired response.

Interacting with large language models (LLM) is no different. The engineer who is interacting with a LLM is offering a prompt, in the form of a request, a desired response has lower a novel discipline: prompt engineering or "the process (and practice) of structuring text that can be interpreted and understood by a generative AI model." [1]

In this article, we explain what is prompt engineering, what constitutes best techniques to engineer a well-structured prompt, and what prompt types are best to LLM to generate the desired response; contrastingly, what prompt types derive an LLM to generate a desired response.

What's Prompt Engineering?

Prompt engineering is not strictly an engineering discipline or precise science, it is more an art than a science, it requires intuition and trial and error, with a set of guidelines to craft precise, concise, creative inquiries or text to instruct an LLM to carry out a task. Another way to put it is that "prompt engineering is the art of communicating with a generative large language model." [2]

To communicate with LLM with precise and task-specific instructions, prompts can range from a single sentence/question to multiple instructions and guidelines to improve the quality of prompts for LLMs. The study suggests that because LLMs exhibit an impressive ability and quality in understanding natural language and in carrying out tasks in various domains such as answering questions, mathematical reasoning, code generation, translation and communication of tasks, etc., principled approach to create concise prompt types and techniques immensely improves a generated response.

Take this simple example of a prompt before and after applying one of the principles of conciseness and precision in the prompt. The difference in the response suggests both readability and simplicity: and the desired target audience for the response.

Prompt: Write a description for a drone product.
Prompt: Write a description for a drone product that targets first time buyers. Describe its ease of use, portability, and its price point.
Llama
Llama 2 is a family of state-of-the-art open-access large language models released by Meta. We're taking a comprehensive approach to scaling Llama 2 to the benefit of developers and Meta alike. This includes a commercial license, a responsible use guide, and a model evaluation and fine-tuning recipe to get the community started on building upon Llama 2. Llama 2 pretrained models are trained on 2 trillion tokens, and have double the context length than Llama 1. Its fine-tuned models have been trained on over 1 million human annotations.
Llama 2 is intended for commercial and research use in English. Its tuned models are intended for assistant-like chat, whereas pretrained models can be adapted for a variety of natural language generation tasks.
Llama 2 pretrained models are trained on 2 trillion tokens of data from publicly available sources. The fine-tuned models were trained for adding chat-like functionality, namely, following instructions. For this, we used publicly available instruction datasets, and over 1 million new human annotations. We used a new mix of publicly available online data. We also a new mix of publicly available online data. We also propose a new technique called Ghost Attention (GAT) to better control dialog flow over multi-turn dialog. We also a new technique called Ghost Attention (GAT) to better control dialog flow over multi-turn dialog.
Llama 2 pretrained models are trained on 2 trillion tokens of data from publicly available sources. The fine-tuned models were trained for adding chat-like functionality, namely, following instructions. For this, we used publicly available instruction datasets, and over 1 million new human annotations. We used

More from Jules S. Damji and The Modern Scientist

In The Modern Scientist by Jules S. Damji

An Exploratory Tour of DSPy: A Framework for Programing...

Is declarative approach appropriate to programming LLMs? A peek at DSPy...

Jun 3, 2024 459 6

In The Modern Scientist by janmeskens

Data Ingestion — Part 1: Architectural Patterns

Over the course of two articles, I will thoroughly explore data ingestion, a...

Nov 27, 2023 2.8K 25

In The Modern Scientist by Yule Wang, PhD

Step-by-Step Illustrated Explanations of Transformer

with detailed explained Attention Mechanism

Feb 27, 2023 416 4

Jules S. Damji

Apache Spark™ 4.0's Variant Data Type Eliminates Complex JSON...

Three synthetic real-world use cases demonstrate gains and usage through binar...

Aug 26 10 2

See all from Jules S. Damji

See all from The Modern Scientist

Recommended from Medium

Sami Kohvakka

How to use Web Search in VS Code Github Copilot

Using MCP tools and Tavily / Bing API key

Apr 15 1

In Data And Beyond by Mangesh Salunke

Prompts in LangChain with Examples

Static, Dynamic and Few-Shot Prompt Templates. Designing Reusable Prompt...

Apr 6 154

In Towards Dev by Saurabh Singh

The Art of Writing Great System Prompts

A Comprehensive Guide for Tech Builders and Decision Makers

Jul 26 4

Cristina CAVALERI

Building an LLM Agent in Python with LangGraph, Mistral, and Tavi...

LangChain, LangGraph, Mistral, Tavily... these names might sound familiar if you've been...

May 12 4

In Level Up Coding by Fareed Khan

Building 17 Agentic AI Patterns and Their Role in Large-Scale AI...

Ensembling, Meta-Control, ToT, Reflexive, PEV and more

5d ago 1.4K 33

Anil Kumar Bakkashetti

10 AI Tools to Double Your Productivity in 2026

Work Smarter, Not Harder Next Year

Sep 23

See more recommendations