University of BRISTOL

LARGE LANGUAGE MODEL AND PROMPT ENGINEERING

BRMS PGR Python Training

Zhaozhen Xu, zhaozhen.xu [at] bristol.ac.uk

11.11.2024

bristol.ac.uk

Isambard-AI

https://www.bristol.ac.uk/news/2023/november/supercomputer-announcement.html

Training large language models (LLMs), big data and robotics

Unprecedented £225m investment to create UK's most powerful supercomputer in Bristol

Press release issued: 1 November 2023

Details of a £225m investment from the Government to create the UK's fastest supercomputer at the University of Bristol have been announced today [1 November 2023].

The funding injection, part of a £300m package to create a new national Artificial Intelligence Research Resource (AIRR) for the country announced at the government's AI Safety Summit at Bletchley Park, will make the UK a world leader in AI.

The National Composites Centre (NCC) which will host Isambard-AI
Image credit: National Composites Centre (NCC)

bristol.ac.uk

How does the Computer Understand text?

Front Cell Infect Microbiol. 2022; 12: 892508. PMCID: PMC9162155
Published online 2022 May 19. doi: 10.3389/fcimb.2022.892508 PMID: 35663468

Effects of COVID-19 Non-Pharmacological Interventions on Dengue Infection: A Systematic Review and Meta-Analysis

Qin Wu,[1,2,3] Shuwen Dong,[1,2,3] Xiaokang Li,[1,2,3] Boyang Yi,[1,2,3] Huan Hu,[1,2,3] Zhongmin Guo,[✉4,*] and Jiahai Lu[✉1,2,3,5,6,*]

Associated Data

Supplementary Materials

Abstract Go to:

Non-pharmacological interventions (NPIs) implemented during the coronavirus disease 2019 (COVID-19) pandemic have demonstrated significant positive effects on other communicable diseases. Nevertheless, the response for dengue fever has been mixed. To illustrate the real implications of NPIs on dengue transmission and to determine the effective measures for preventing and controlling dengue, we performed a systematic

Language Modelling!

A model learns to understand and predict text by capturing patterns and relationships between words.

Large Language Model

What is a Large Language Model (LLM)?

A model that can understand and generate human-like text by:

- complex neural network architectures
  - Transformer: millions to billions parameters
  - BERT based: 110M parameters
  - Llama 2: 70B parameters
- reading almost everything on the Internet
  - Pre-training

Most of the current LLMs are built using **Transformer** architecture.

Encoder

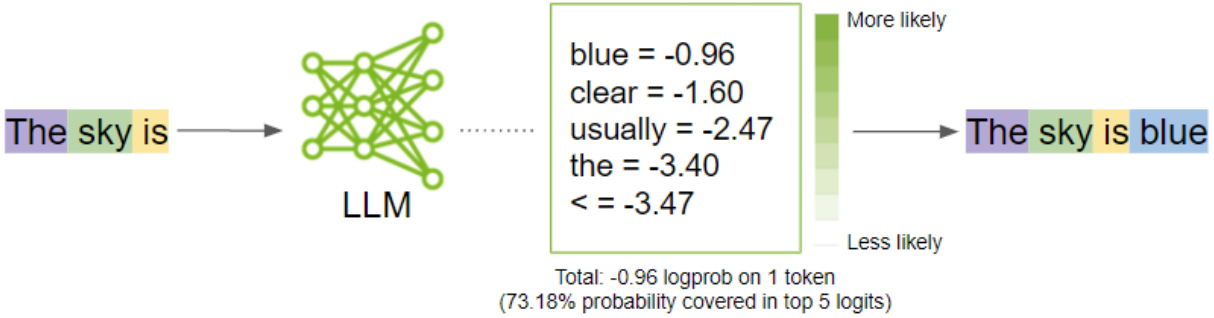
- Map the input to a fixed-size context vector

Decoder

- Convert the context vector back to text

Source: 'Attention is all you need' (Vaswani et al. 2017)

How does Large Language Model obtain knowledge



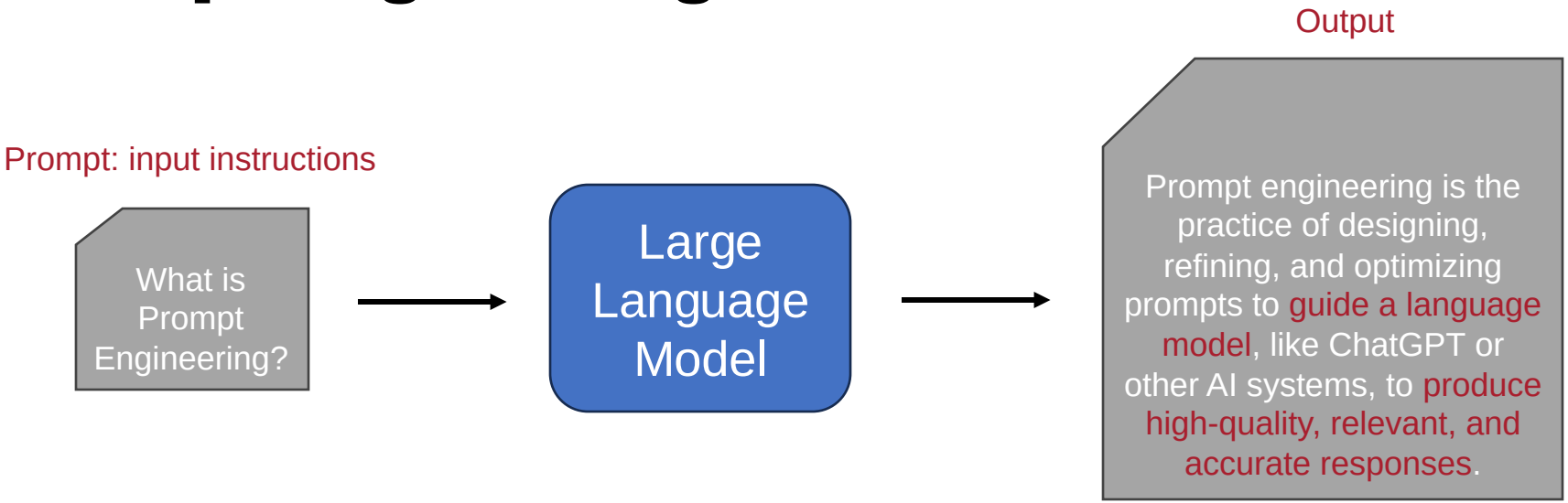
Source: https://developer.nvidia.com/blog/how-to-get-better-outputs-from-your-large-language-model/

- The primary goal of language modelling is to predict the next word in a sequence of text given the preceding words.
  - Understand the structure, grammar, context, and meaning in language
- Pre-trained on Large Amount of Text
  - Book, Websites, Newspaper, …
  - Domain Specific Knowledge: PubMed, MIMIC (Medical Information Mart for Intensive Care)

Large Language Model in Health Science

- **Medical Documentation and Summarization**
  - Generate structured summaries (e.g. diagnoses, treatment plans) from patient interviews
- **Support Systematic Literature Review**
  - Automating Study Screening, Data Extraction, Risk of Bias Analysis
- **Drug Discovery**
  - Analyse the reasons for stopping trials
- **Electronic Health Record (EHR) Analysis**
- **Chatbots for Patient Support**
- **Medical Image Analysis**
  - combined with other computer vision models

Prompt Engineering



- Zero-shot: ask the LLM to perform a task without examples
- N-shot: a prompt includes N examples

Prompt Structure

- System Prompts
  - A special instruction given to the model at the start of a conversation
  - Setting the overall tone, role, and boundaries for the interaction

- User Prompts (i.e. the main prompt)
  - The main input provided by the user
  - Can be a question, instruction, or request for information

- Example Prompts (seen in N-shot prompting)
  - input and output examples provided within the prompt
  - help the model understand the structure, style, or specifics of the desired response.
  - helpful for more complex tasks.

Example

- Sending prompts in batch
- Getting response from Llama 2 using Python and HuggingFace

- Acknowledgement: The code is provided by Olalekan Awoniran.

Reference

- COSTAR framework: https://medium.com/@frugalzentennial/unlocking-the-power-of-costar-prompt-engineering-a-guide-and-example-on-converting-goals-into-dc5751ce9875