Software Requirements Engineering (SRE)

through

Large Language Models (LLMs)

Pir Sami Ullah Shah

What are LLMs?

What is LLM?

“Refers to an Artificial Intelligence (AI) model that has been trained on large amounts of data and is able to generate text in a human-like fashion.”

What is LLM? (CONTD)

- LLMs are typically based on deep learning techniques, such as transformers.
- Capable to generate useful language output. As a result, they have been found capable of performing a wide range of language-related tasks, including
  - Text generation.
  - Answering questions.
  - Translation.
  - Summarization.
  - Sentiment Analysis.

LLM Back Tracking

- Rumelhart et al. introduced the concept of **Recurrent Neural Network**, opening up the possibility of processing sequential data.
- **Long Short Term Memory (LSTM)** architecture, an extension of the RNN architecture introduced by Hochreiter and Schmidhuber, significantly improved their performance in many applications.

LLM Back Tracking (CONTD)

- In 2017, Vaswani et al. introduced the **Transformer** architecture, which captures word relationships with the self attention mechanism.
- The transformer architecture had a profound impact on language modelling and triggered an explosion of activity on LLMs.

LLM Back Tracking (CONTD)

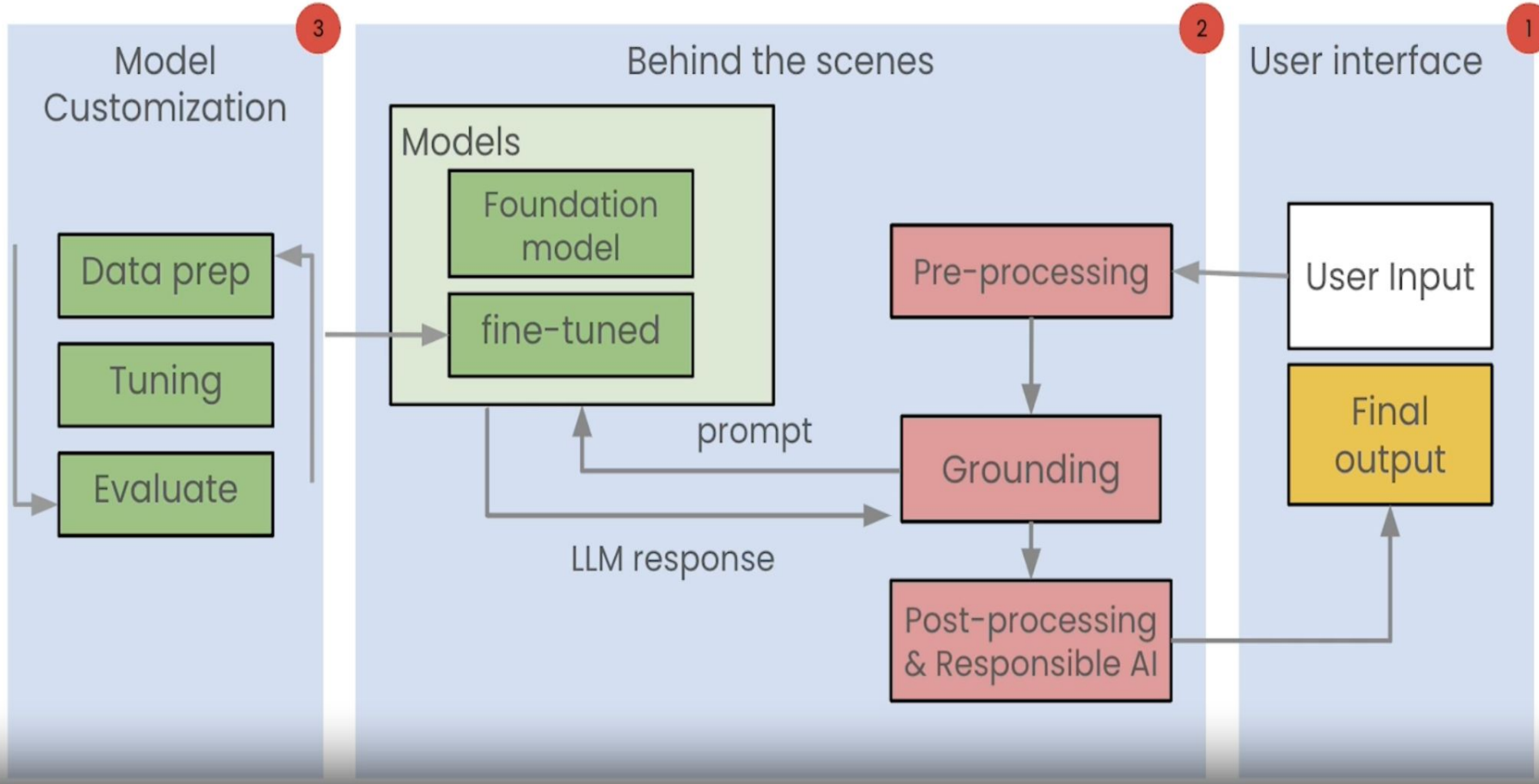
- In 2018, OpenAI released the Generative Pre-trained Transformer (GPT) model, followed by subsequent iterations (GPT2, GPT-3, GPT-3.5, and GPT-4).
- With GPT-3 and 3.5, many observers noticed a significant step change in generative performance, thereby attracting a great deal of interest in GPT (and ChatGPT) in particular, and also in LLMs more generally.

Any particular reason behind the surge of LLMs in almost every domain?

Reasons behind the surge

- Due to the large corpora (dataset) on which they (LLMs) are trained on.
- GPT-3 is trained on 45TB of text data and has 175 billion parameters.
- Meta launched open-sourced LLaMA in February 2023, which is trained on 1.4 trillion tokens with a variety of model sizes ranging from 7 billion to 65 billion parameters.

High level example of a LLM driven application



Name	Release date	Produced by	Parameters	Open-sourced	Price	Supported languages
CodeBERT	February 2020	Microsoft	125M	YES	free	6
InCoder	April 2022	Meta	6.7B, 1.3B	YES	free	30
AlphaCode	February 2022	DeepMind	300M, 1B, 3B, 9B, and 41B	NO	free	Python or C++
CodeX	August 2021	OpenAI	12B	NO	free	>11
Copilot	October 2021	Github and OpenAI	12B	NO	free for individual developers and organisations	>11
CodeT5	Nov 2021	Salesforce Research	60M, 220M, 770M	YES	free	6
CodeT5+	May 2023	Salesforce Research	2B, 6B, 16B	YES	free	9
PolyCoder	Oct 2022	Carnegie Mellon University	160M, 400M, 2.7B	YES	free	>11
CodeWhisperer	April 2023	Amazon	unknown	NO	free for individual developers	15
WizardCoder	June 2023	Microsoft	15B	YES	free	unknown
CodeGeeX	Sep 2022	Tsinghua University et al.	13B	YES	free	23
CodeGen	March 2022	Salesforce Research	350M, 1B, 3B, 7B, 16B	YES	free	Python
StarCoder	May 2023	BigCode	15B	YES	free	>80
phi-1	June 2023	Microsoft	1.3B	NOT YET	free	Python
Code Llama	August 2023	Meta	7B, 13B, 34B	YES	free	>7

Software Requirements Engineering

Why are Requirements more important?

How LLMs can help SRE?

Point 1

- LLMs can help requirement engineers to highlight the unknowns in requirements documentations.
- Unknowns could be any ambiguity or uncertainty in requirements.
- Can help with the completion or suggest alternative ideas regarding the identified unknowns.

Point 2

- Classification of requirements into Functional (FR)or Non-Functional Requirement (NFR).
- Identification of FRs and NFRs from user reviews on platform like play store or app store.

Point 3

- Extraction of user stories from a single requirement provided by the end user.

Point 4

- Extraction of user stories from a single requirement provided by the end user
- Converting raw requirements into well constructed and structured documentations.

Point 5

- LLMs can also suggest requirements prioritization by analyzing technical dependencies, project goals, and historical data.

Conclusion

- One can clearly see potentials of LLMs in software requirements engineering.
- There is very limited literature available to assist SRE by utilizing LLMs.
- The court is open and you (students) are supposed to drop a ball in the open court.

Contact Details

Pir Sami Ullah Shah,

Lecturer, Department of Software Engineering,

NUCES, Islamabad.

Office: A304F

Office Hours: 0200 (pm) onwards (Monday and Wednesday)

samiullah.shah@nu.edu.pk