

Leveraging Large Language Models in Software Engineering

Techniques, Challenges, and Opportunities

Gianluca Aguzzi gianluca.aguzzi@unibo.it

Dipartimento di Informatica – Scienza e Ingegneria (DISI)
Alma Mater Studiorum – Università di Bologna

01/03/2024



Outline

- 1 Introduction to LLMs
- 2 Machine Learning for Software Engineering
- 3 LLMs in Software Engineering
- 4 LLMs for Software Testing
- 5 Innovative Solutions for LLMs – autonomous agent models

Natural Language Processing (NLP)



Natural Language Processing (NLP)

A subfield of artificial intelligence that focuses *understanding*, *interpreting*, and *generating* human language.



Natural Language Processing (NLP)

A subfield of artificial intelligence that focuses *understanding*, *interpreting*, and *generating* human language.

Resources

- <https://github.com/keon/awesome-nlp?tab=readme-ov-file>
- <https://github.com/brianspiering/awesome-dl4nlp>
- <https://nlpprogress.com/>
- <https://www.unibo.it/it/studiare/dottorati-master-specializzazioni-e-altra-formazione/insegnamenti/insegnamento/2023/412644>



Natural Language Processing

Goal

Identify the structure and meaning of *words*, *phases*, and *sentences* in order to enable computers to understand and generate human language.

Why?

Improve *human-computer* interaction, closing the gap between *human communication* and *computer understanding*.

Applications (all around us)

- *Chatbots*
- *Machine Translation*
- *Speech Recognition*
- *Sentiment Analysis*
- *Question Answering*
- *Code Generation*

Natural Language Processing

Challenges

- **Ambiguity:** Multiple meanings for words/phrases.
- **Context:** Meaning shifts with context (linguistic, cultural).
- **Syntax:** Sentence structure affects meaning.
- **Sarcasm/Idioms:** Non-literal language interpretation.

Approaches

- **Rule-Based:** Hand-crafted linguistic rules (e.g., Georgetown-IBM).
 - **Statistical:** Probabilistic language modelling (e.g., hidden Markov model [1]).
 - **ML/Deep Learning:** Algorithms learn from data; neural networks model complex patterns (RNN, LSTM, etc.).
- ➔ we will focus on ***Language Models***.

What is a **Language Model**?



What is a **Language Model**?

A *machine learning* model that aims to predict and generate plausible text.



Language Models

Idea

text is a sequence of words, and language models learn the *probability* of a word given the previous words.

Example

- *The cat is on the <*>*
- *The cat is on the **mat**.*
- *The cat is on the **table**.*

Tasks

- *Prediction*: Given a sequence of words, predict the next word (e.g., autocomplete).
- *Classification*: Given a sequence of words, classify the text (e.g., sentiment analysis).

Leveraging Large Language Models in Software Engineering

Techniques, Challenges, and Opportunities

Gianluca Aguzzi gianluca.aguzzi@unibo.it

Dipartimento di Informatica – Scienza e Ingegneria (DISI)
Alma Mater Studiorum – Università di Bologna

01/03/2024



References I

- [1] Bernard Mérialdo. “Tagging English Text with a Probabilistic Model”. In: *Comput. Linguistics* 20.2 (1994), pp. 155–171.

