Intelligenza Artificiale



An introduction to NLP

ALMA MATER STUDIORUM Università di Bologna

> Ivan Heibi Dipartimento di Filologia Classica e Italianistica (FICLIT) <u>Ivan.heibi2@unibo.it</u>

Speech and Language Processing

An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition

Third Edition draft

Daniel Jurafsky Stanford University

James H. Martin University of Colorado at Boulder

https://web.stanford.edu/~jurafsky/slp3/

Introduction to NLP

People communicate using language, whether it's written or spoken

For an effective interaction between computers and humans, it's crucial for computers to comprehend the natural languages that humans use.

Natural Language Processing (NLP) is dedicated to teaching computers how to <u>learn, process, and manipulate</u> <u>human languages</u>



NLP practices

NLP is used to comprehend the structure and significance of human language through the examination of various components such as syntax, semantics, pragmatics, and morphology.

Macro categories are:

- **Syntactic analysis** (i.e., parsing or syntax analysis): recognizing the syntactic structure within a text and the dependency relationships among words.
- **Semantic analysis:** identifying the meaning of language. (since language is polysemic and ambiguous, semantics is considered one of the most challenging areas in NLP)

Some relevant sub-tasks:

- Word Sense Disambiguation
- Text Classification (e.g., Topic Modeling, Sentiment analysis)
- Named Entity Recognition

NLP processing pipeline



Large Language Models (LLMs)

Large language models (LLMs) are massive neural networks trained on enormous amounts of text data (i.e., hundreds of billions of words). ChatGPT and Gemini are Large language models.

LLMs can comprehend long-term dependencies, intricate relationships among words, and subtleties inherent in natural language. LLMs have the capability to process all words simultaneously, leading to accelerated training and inference.

<u>Transformer models (Deep learning technique) is used during the modelling step (i.e.</u> <u>step 4) of the NLP pipeline</u>

Transformers are specifically designed to capture relationships in sequential data, relying on a self-attention mechanism to grasp global dependencies between input and output.