

Natural Language Processing

Natural Language Processing

1. Introduction

- **Natural Language Processing (NLP):** a branch of AI concerned with enabling computers to understand, interpret and generate human language.
- It bridges linguistics and computer science to process text and speech.
- Underpins translation, chatbots, search engines and voice assistants.

2. Core Tasks

- **Tokenisation:** splitting text into words or sub-word units.
- **Part-of-speech tagging:** labelling words by grammatical role.
- **Named entity recognition:** identifying names, places and dates.
- **Sentiment analysis:** detecting opinion or emotion in text.

3. Key Techniques

- Rule-based and statistical methods preceded modern deep learning.
- **Word embeddings:** vector representations capturing word meaning and similarity.
- **Transformers:** attention-based architectures (e.g. BERT, GPT) that power large language models.

4. Applications

- Machine translation, text summarisation and question answering.
- Spam filtering, autocomplete and grammar correction.
- Conversational agents and voice-controlled assistants.

5. Challenges

- Ambiguity, sarcasm, idioms and context dependence.
- Handling multiple languages and low-resource dialects.
- Bias in data and the high cost of training large models.