Natural Language Processing (NLP)

Artificial Intelligence Applications

(Case-studies)

Eng. Mustafa Othman
Data Scientist & Analyst

Full Course Outline:

- **NLP Basics**
  - What? & Why?

- **NLP Applications**
  - NLP Tasks

- **NLP Pipeline**

- **NLP Toolkit**

Natural Language Processing (NLP) Basics

"A different language is a different vision of life." ~ Federico Fellini.
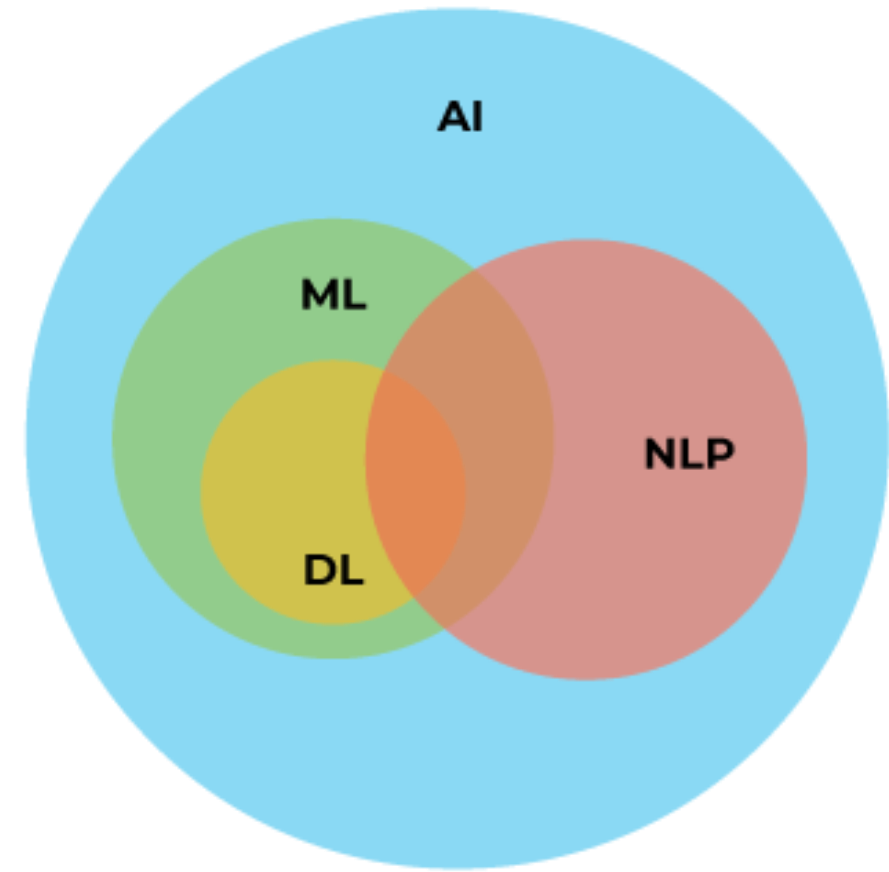
Eng. Mustafa Othman
Data Scientist & Analyst

AI will probably kill us!

- Microsoft's **Tay** AI Chatbot (2016)
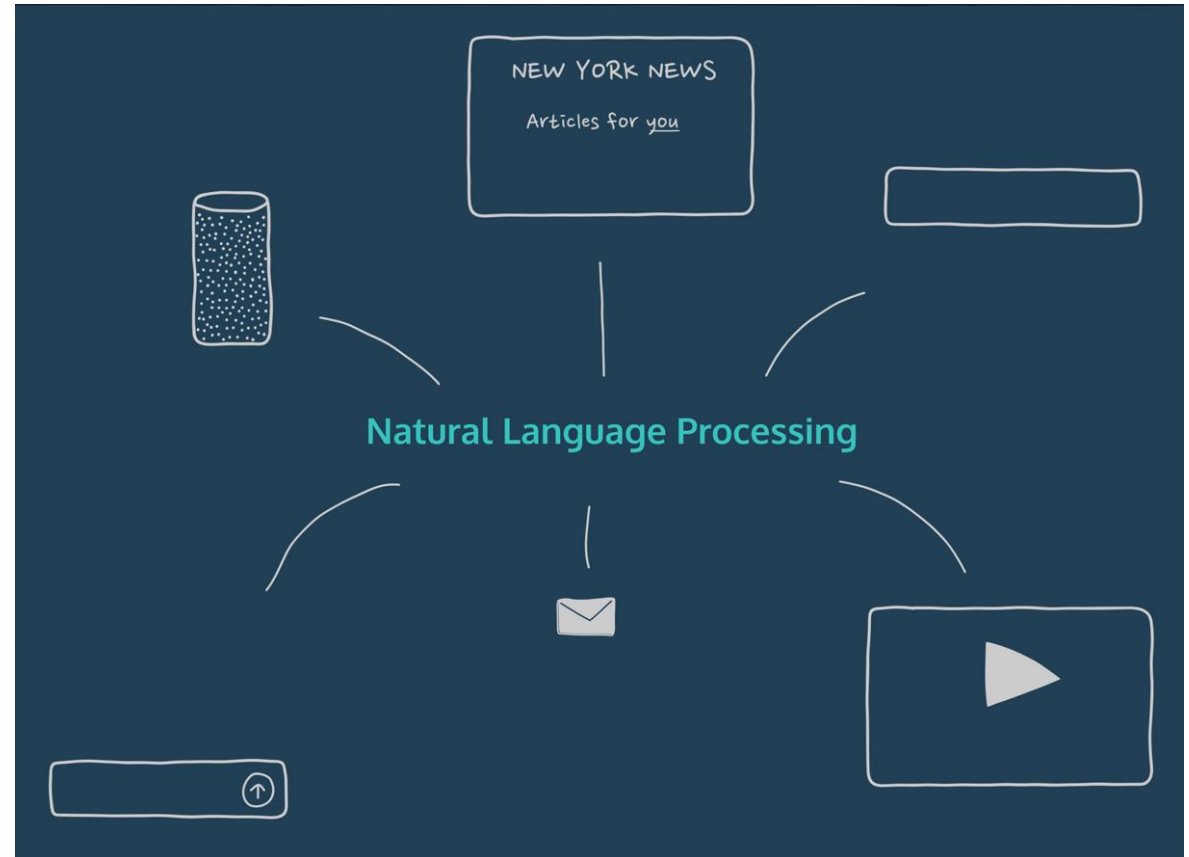 - https://youtu.be/jTSn7f4sEKo

Natural Language Processing (NLP) Basics

- **Why?**
  - Huge Amount of Textual Data : Email, SMS, Websites, .. Etc.
  - Challenges: Ambiguity, Creativity, Human Knowledge, .. Etc.
- **What?**
  - **Natural language processing (NLP)** is the ability of a computer program to understand **human language** as it is spoken and written.
- **NLP Components:**
  - Natural Language Understanding (**NLU**)
  - Natural Language Generation (**NLG**)
- **Approaches to NLP**
  - Statistical Methods
  - Neural Networks

Eng. Mustafa Othman
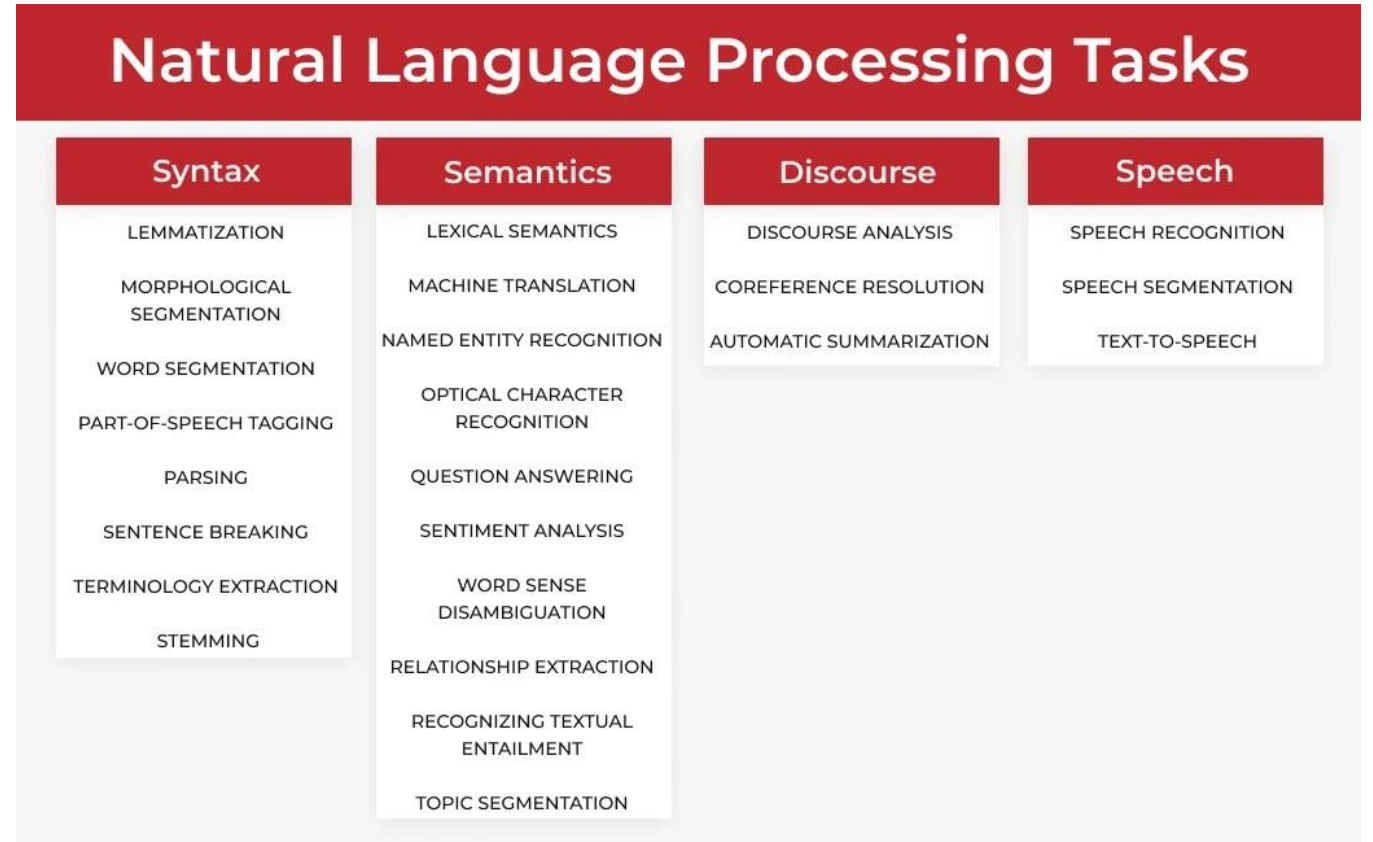Data Scientist & Analyst

NLP Applications

- Text Classification
- Sentiment Analysis
- Machine Translation
- Virtual Agents & Chatbots
- Text Summarization
- Question Answering System
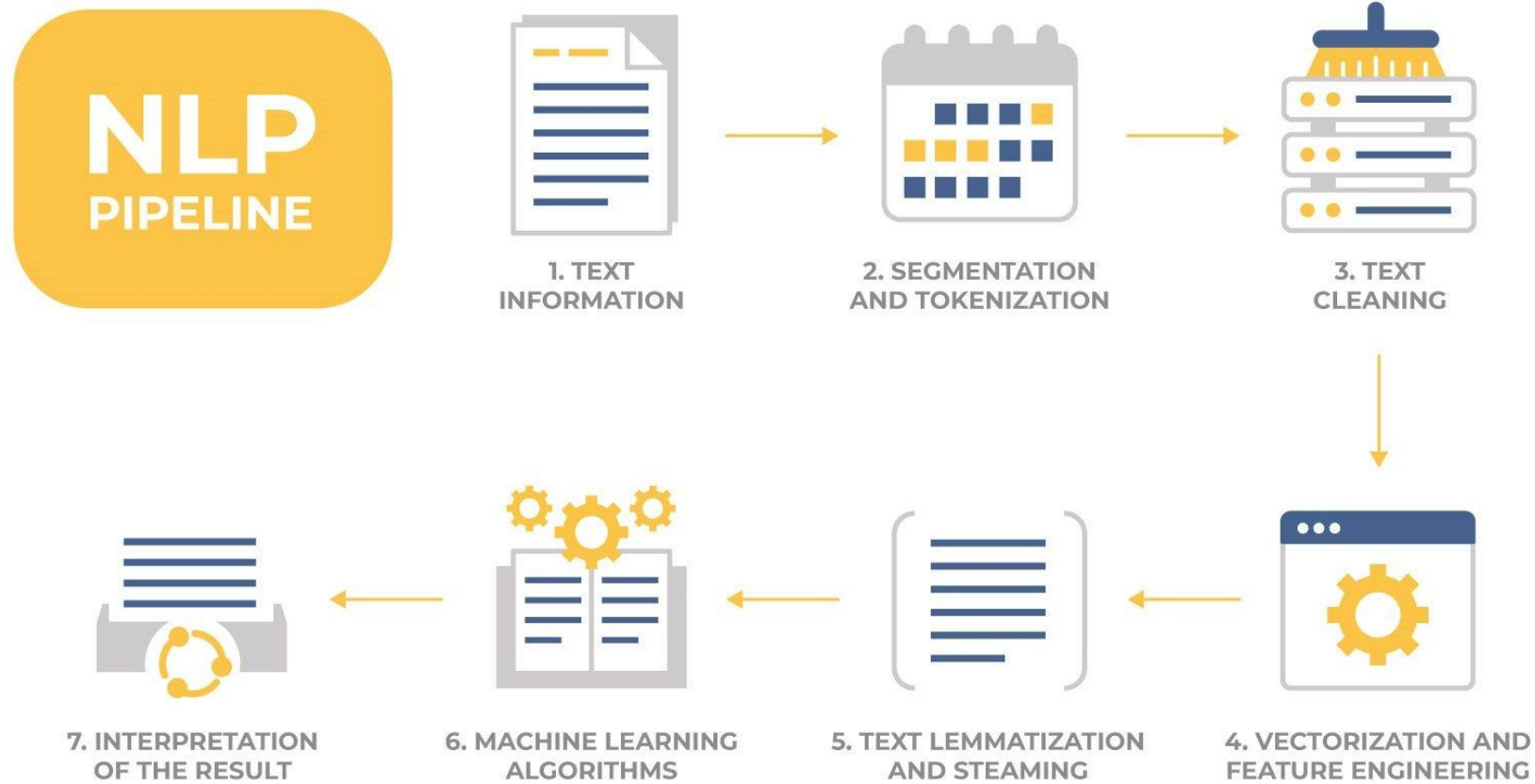- Information Retrieval & Extraction
- Spam Detection

NLP Tasks

- Tokenization
- Stop Words Removal
- Lemmatization
- Stemming
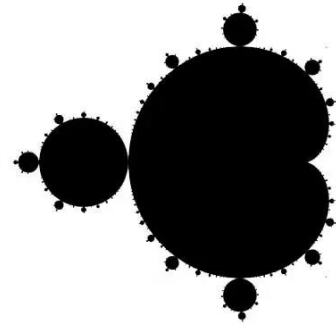- Part-of-Speech (POS) Tagging
- Named Entity Recognition (NER)

Natural Language Processing Tasks

Syntax	Semantics	Discourse	Speech
LEMMATIZATION	LEXICAL SEMANTICS	DISCOURSE ANALYSIS	SPEECH RECOGNITION
MORPHOLOGICAL SEGMENTATION	MACHINE TRANSLATION	COREFERENCE RESOLUTION	SPEECH SEGMENTATION
WORD SEGMENTATION	NAMED ENTITY RECOGNITION	AUTOMATIC SUMMARIZATION	TEXT-TO-SPEECH
PART-OF-SPEECH TAGGING	OPTICAL CHARACTER RECOGNITION		
PARSING	QUESTION ANSWERING		
SENTENCE BREAKING	SENTIMENT ANALYSIS		
TERMINOLOGY EXTRACTION	WORD SENSE DISAMBIGUATION		
STEMMING	RELATIONSHIP EXTRACTION		
	RECOGNIZING TEXTUAL ENTAILMENT		
	TOPIC SEGMENTATION		

Eng. Mustafa Othman
Data Scientist & Analyst

NLP Pipeline

NLP PIPELINE

1. TEXT INFORMATION

2. SEGMENTATION AND TOKENIZATION

3. TEXT CLEANING

4. VECTORIZATION AND FEATURE ENGINEERING

5. TEXT LEMMATIZATION AND STEAMING
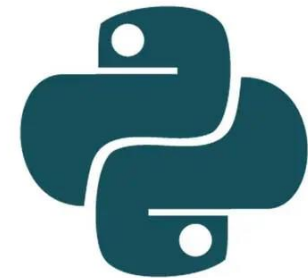
6. MACHINE LEARNING ALGORITHMS

7. INTERPRETATION OF THE RESULT

NLP Toolkit

- NLTK
- spaCy
- TextBlob
- Gensim
- Sci-kit Learn
- TensorFlow
- PyTorch

THANKS

Keep Moving Forward! ☺

Eng. Mustafa Othman
Data Scientist & Analyst