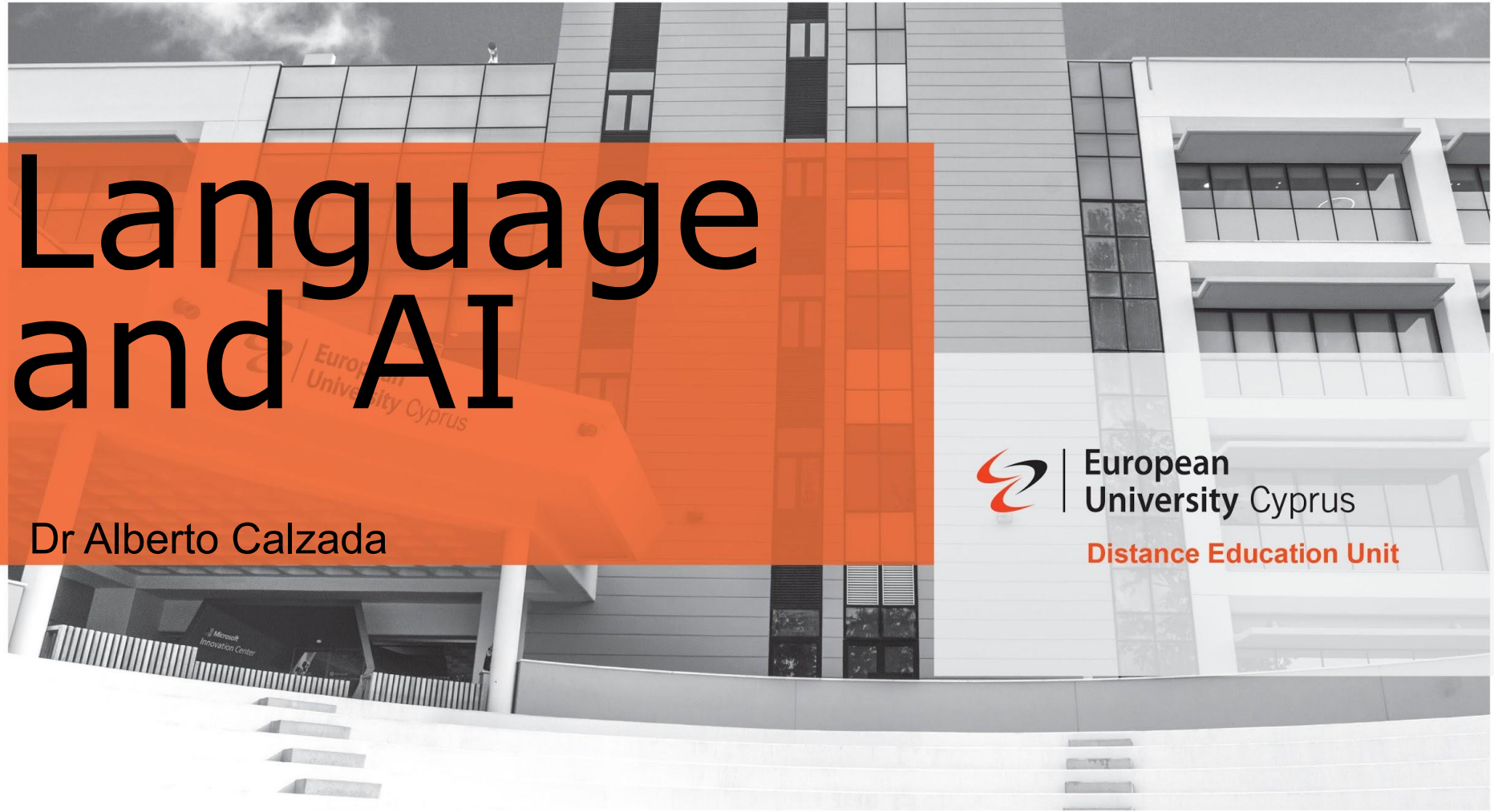
Language and AI

Dr Alberto Calzada



European University Cyprus

Distance Education Unit

Introduction

Alberto Calzada, PhD

Data Science and Engineering Consultant

10+ years industry and academic experience

Lecturer of: Machine Learning, Big Data Analytics, Deep Learning and **Natural Language Processing** at the European University Cyprus.

Coordinator of MSc in AI (Distance Learning programme)

Content

- Introduction
- What is AI? What is Machine Learning?
- Language and AI: Natural Language Processing
- What can we do with NLP? Tasks
- History: where are we heading?

What is AI? What is Machine Learning?

Artificial Intelligence (AI):

An area inside Computer Science that focuses on creating machines able to replicate or imitate intelligent behaviours.

Usually divided into **Weak** and **Strong** AI

When talking about AI in general, we are commonly referring to Weak AI: An algorithm that solves a specific task.

What is AI? What is Machine Learning?

Machine Learning (ML):

A branch of Artificial Intelligence focused on providing machines with learning capabilities.

ML is usually divided into 3 learning paradigms: supervised, unsupervised and reinforced learning

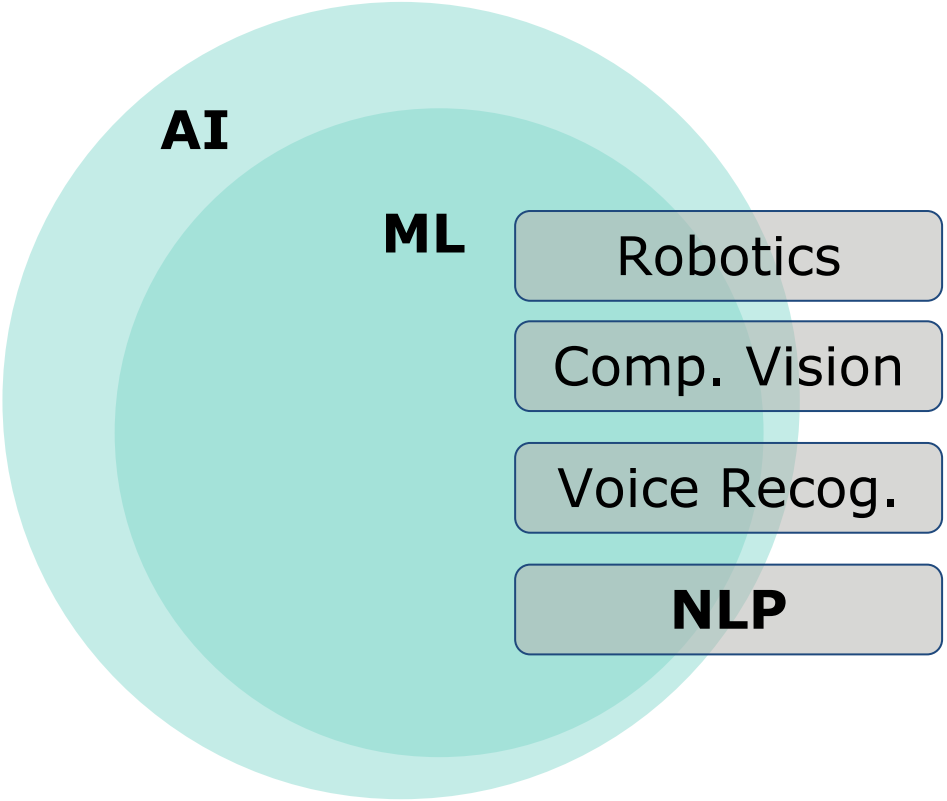
Learning capabilities: In the vast majority of cases, these learning capabilities are acquired from data

What is AI? What is Machine Learning?

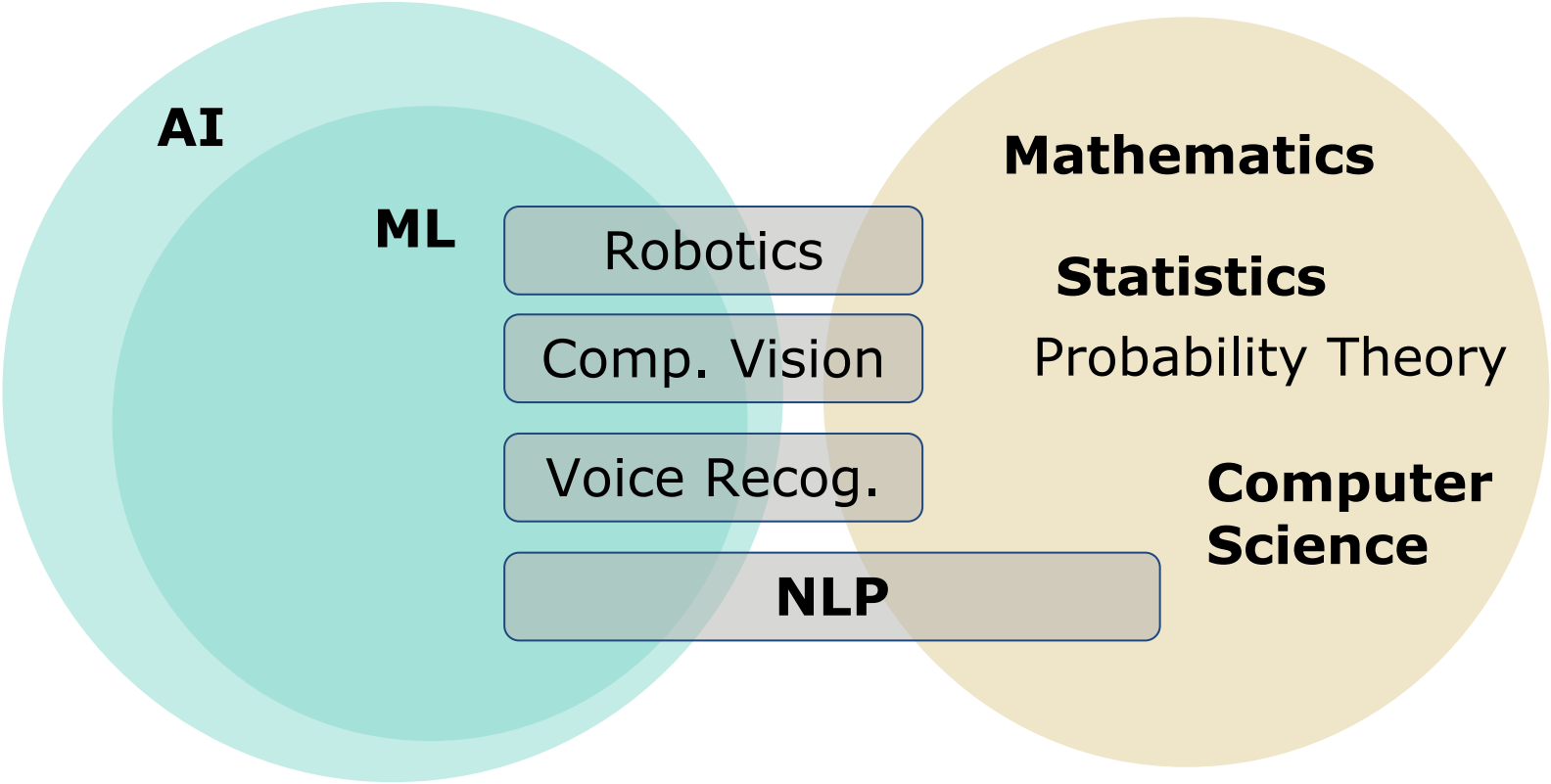
Specialised areas of knowledge and application that take advantage of AI and/or ML:

- Voice Recognition
- Natural Learning Processing (NLP)
- Robotics
- Computer Vision
- etc.

What is AI? What is Machine Learning?



What is AI? What is Machine Learning?



Natural Language Processing (NLP)

Natural Language Processing (NLP):

A collection of techniques focused on the interactions between computers and human language, and how to process, analyse, recognise, classify, extract and generate information from large amounts of natural language data, usually provided as text.

Natural Language Processing (NLP)

Natural Language Processing (NLP):

*A collection of techniques focused on the interactions between computers and human language, and how to **process**, **analyse**, **recognise**, **classify**, **extract** and **generate** information from large amounts of natural language data, usually provided as text.*

NLP: A task-centered field of research and application

NLP Tasks

Low-level tasks: Simpler tasks such as

- extracting certain word patterns from text
- identify the lemmas or basic concepts
- recognise named entities
- create language models

These low-level tasks can be used for spell-checking, basic information retrieval or next-word suggestion.

They usually need one document to work with.

NLP Tasks

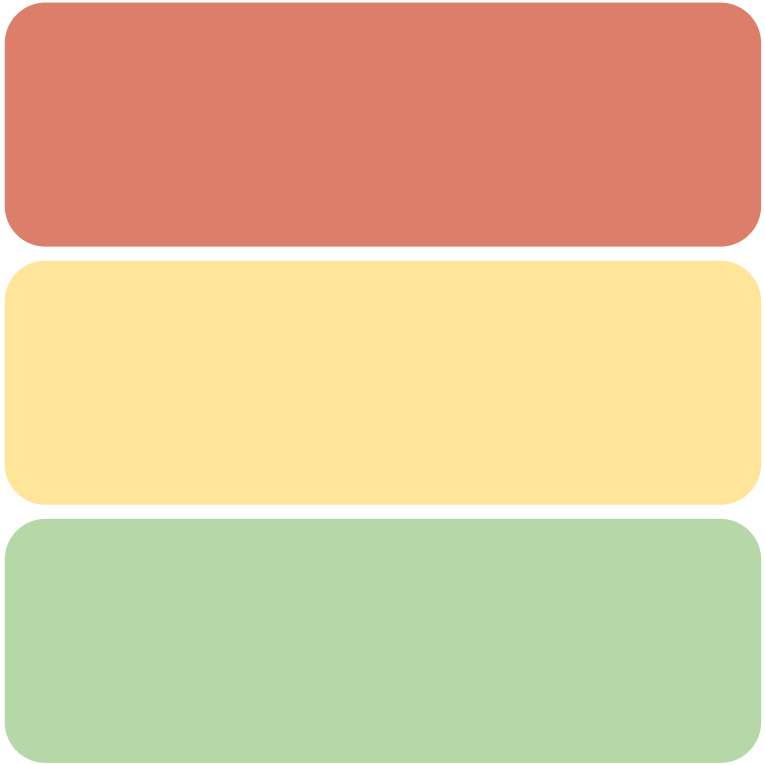
Mid-level tasks: Increased complexity

- analyse the topic of a text
- classify documents based on their content
- extract information about named entities
- create embeddings

The use of these mid-level tasks is widespread in industry.

They usually require many documents (a corpus) - the more documents, the better they will work.

NLP Tasks

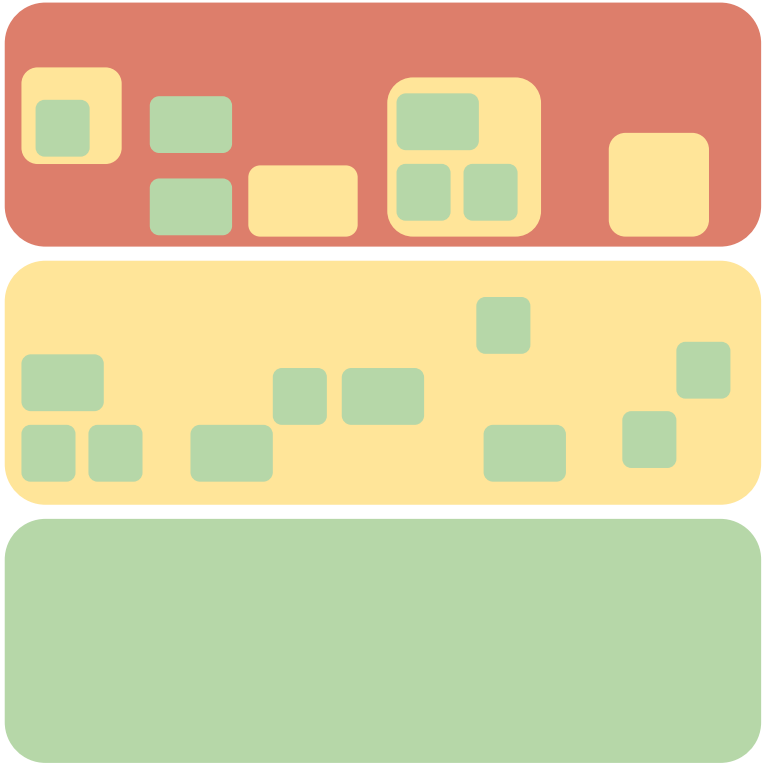


High-level tasks: Complex problems

- automated text summarisation
- machine translation
- chatbots

Starting to be used in industry, mainly by large tech corporations. Their design is either a composite of low and mid-level tasks, a Deep Learning model (usually a Transformer) or a combination of both.

NLP Tasks



Low and mid-level tasks are in many cases the base for more complex tasks.

In this way, NLP can be regarded as a collection of “building blocks” which we can combine to create more complex systems.

History

Procedural Era
1950s-1960s

- Hand-written programmes
- Rule-based Systems: IF-THEN
- First Chatbots

Statistical Era
1970s-1990s

- Language Models
- Probabilistic Approaches
- Bayesian Approaches
- Ontologies

ML Era
2000s-

- Document Classification
- Topic Extraction
- Word Embeddings

Deep Learning Era
2020s-

- RNNs
- Transformers
  - BERT (Google)
  - GPT (Microsoft)

History

Procedural Era
1950s-1960s

- Hand-written programmes
- Rule-based Systems: IF-THEN
- First Chatbots

Statistical Era
1970s-1990s

- Language Models
- Probabilistic Approaches
- Bayesian Approaches
- Ontologies

ML Era
2000s-

- **Document Classification**
- Topic Extraction
- **Word Embeddings**

Deep Learning Era
2020s-

- RNNs
- **Transformers**
  - **BERT (Google)**
  - **GPT (Microsoft)**

Document Classification

To automatically apply a category to a document, given historic training data:

- Spam / No Spam,
- Car Accident Insurance Claim / Medical Insurance Claim
- Sentiment Analysis: Positive/Negative film reviews, Tweets, product reviews…

Probabilistic Models (Naive Bayes) has proven to work very well in industrial applications

	Cat	Documents
Training	-	just plain boring
	-	entirely predictable and lacks energy
	-	no surprises and very few laughs
	+	very powerful
	+	the most fun film of the summer
Test	?	predictable ~~with~~ no fun

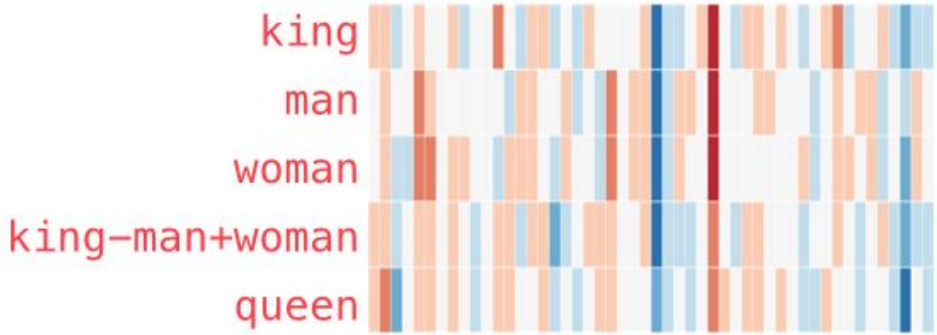
"Speech and Language Processing: An Introduction to NLP, computational linguistics and speech recognition", Jurafsky D. and Martin J.H. (2008), Prentice Hall

Word Embeddings

Abstract, high-level representations of the concepts being represented by words. Such representations are basically vectors of several hundreds/thousands of decimal values, typically between 0 and 1.

- Obtained via Artificial Neural Networks using large corpora (the whole Wikipedia, publications and text of any type, etc).
- Allow us to perform operations on the words, and measure their similarity, to extract information even when an exact keyword is not present in the text

king − man + woman ~= queen



Demo from Google

Transformers

Transformers are used to solve many tasks, such as: machine translation, automatic text summarisation, text generation, named-entities recognition, and analysis of DNA sequences, among others (yes, I know DNA sequences are not NLP, but DNA are sequences of genes, and text are sequences of words - and for our computers these are pretty much equivalent).

Just to give you an idea, here is a diagram of BERT's structure. It is composed of 12 successive transformer layers, each having 12 attention heads.

The total number of hyperparameters that had to be tuned for this pre-trained model is 110 million.