

AI IN VISION

NATURAL LANGUAGE PROCESSING

AI CLUB WEEK #4

PRESENTED BY MICHAL, AREN, SERGI

AI CLUB'S ROADMAP



WEEK 1
Fundamentals

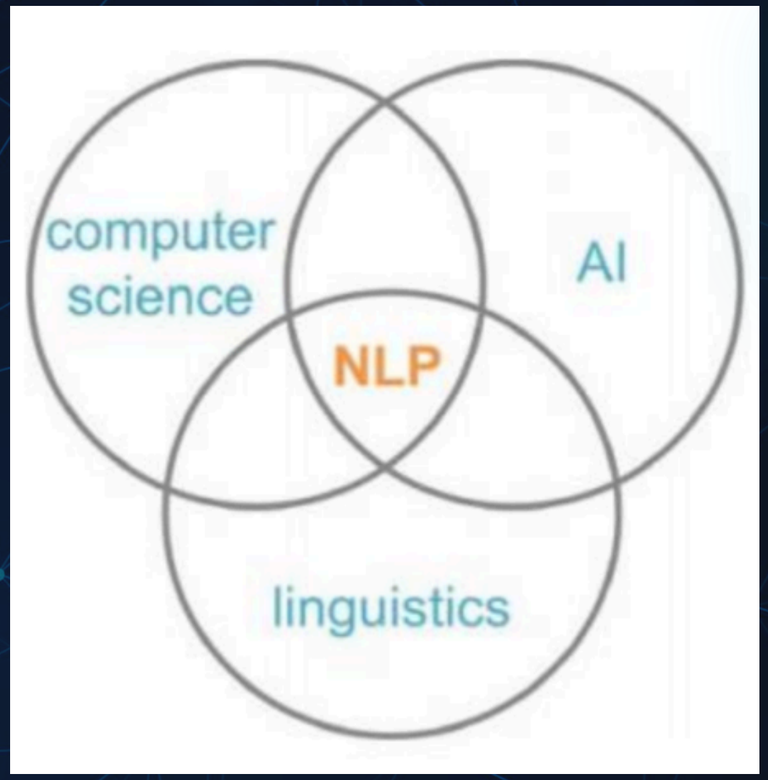
WEEK 2
LLMs

WEEK 3
Diffusion & Agents

WEEK 4
NLP & Vision

WEEK 5
Ethics & Future

WEEK 6+
Classifier Models



WHAT IS NATURAL LANGUAGE PROCESSING?

NLP is the **branch of AI** that gives computers the ability to **understand**, **interpret**, and **manipulate** human language.

With the goal to bridge **ambiguous** human language and **logical** computer understanding.

DIFFERENTIATING LLMS & NLP

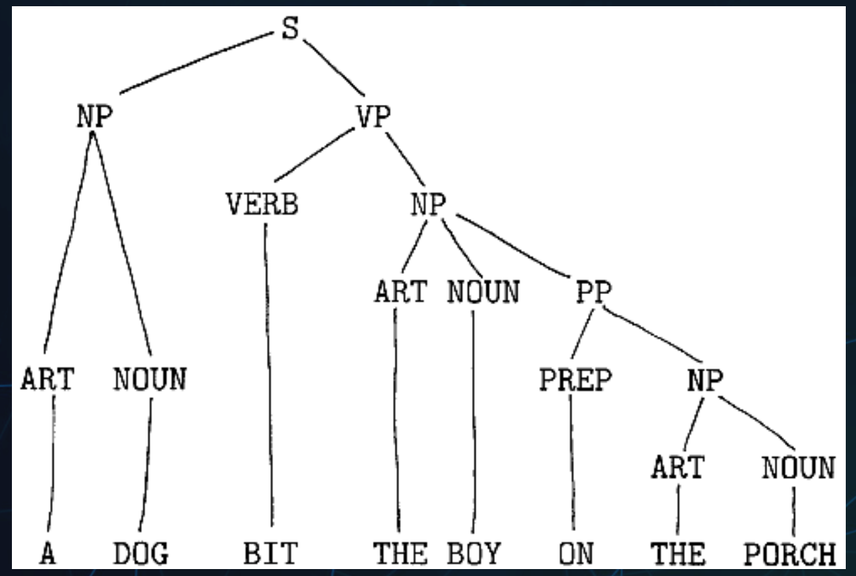
LLMS

Focused on generating text based on probability.

NLP

The broader field of analyzing structure, sentiment, and meaning.

HOW COMPUTERS "READ" - SYNTAX VS. SEMANTICS



SYNTAX

Arranging words to make grammatical sense.

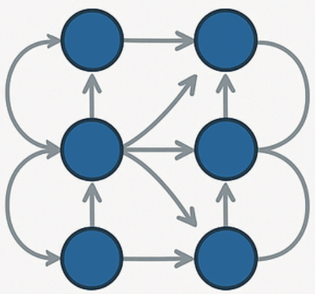
Tokenization: Breaking text into units (words/sub-words).
Lemmatization: Reducing words to their root.

SEMANTICS

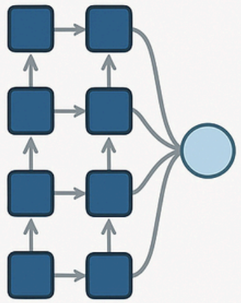
Understanding the logic and definition behind the words.

Named Entity Recognition (NER): Identifying "London" as a location, not just a noun.
Disambiguation: Determining if "Bank" means a river side or a financial building.

THE EVOLUTION - RNNS VS. TRANSFORMERS



RNN



Transformer

THE OLD WAYS: RECURRENT NEURAL NETWORKS

Processes data **sequentially** (word by word).

Problem: Forgets the beginning of a long sentence by the time it reaches the end (**Vanishing Gradient**).

THE NEW WAY: TRANSFORMERS

Processes the entire sentence **simultaneously** (Parallel processing).

Attention Mechanism: **Weighs** the importance of words in **relation** to each other, regardless of distance.

SENTIMENT ANALYSIS - PRACTICAL NLP

THE TASK: CLASSIFYING TEXT BASED ON EMOTIONAL TONE.

HOW IT WORKS:

Assigns "**weights**" to words (e.g., "Good" = +1, "Terrible" = -1).

Analyses **context** (e.g., "Not good" flips the polarity).

INDUSTRY USES CASES:

Finance: Analyzing news to predict stock market movement.

Marketing: Monitoring brand reputation on X (Twitter).

THE LIMITATIONS (WHY IT FAILS):

Sarcasm & Irony: A machine sees "Great, another flat tire" as Positive (+0.8) because of the word "Great," missing the context.

Nuance: "The movie was not terrible" is technically positive, but weak. Basic models struggle with negations.

Aspect-Based Sentiment: Real-world data is mixed. "The food was delicious, but the service was rude."

THE CHALLENGE OF AMBIGUITY

WHY NLP IS HARD:

Human language is **rarely literal**.

Syntactic Ambiguity:

"I saw the man with the telescope."
Did I use a telescope to see him? Or did he have a telescope?

CULTURAL CONTEXT:

Idioms like "Break a leg" or "It's raining cats and dogs" **confuse** basic models.

GETTING HANDS-ON - TOOLS & LIBRARIES

Python Libraries (The Industry Standard):

NLTK (Natural Language Toolkit): The academic standard. Perfect for understanding the theory of tokenization and grammar trees, but slow for massive datasets.

spaCy: The industrial standard. Built for speed and used in actual software products. It comes with pre-trained models 'out of the box'.

Hugging Face: The modern standard. Access to state-of-the-art Transformers (BERT, GPT, Llama) with just a few lines of code.

Where to Get Data:

- **Kaggle Datasets:** Search for "The Great Cognitive Reversal" or "AI Impact on Jobs"
- **Common Crawl:** Massive, free datasets of the open web.

Project Idea: Build a simple spam filter using a "Bag of Words" model.

CONCLUSION & NEXT STEPS

SUMMARY:

NLP allows computers to **read**, **decipher**, and **understand** human languages.

It moves **beyond** simple keywords to understanding **intent** and sentiment.

NEXT WEEK:

We will cover **Cybersecurity** and **Ethics**, and will be sending resources ahead of the week such as articles in preparation for this.

Resources for this Week: Everything from previous weeks, courses, articles posted, with recap of previous slides to ensure preparation to delve into more detail next week.

THANK YOU

If you are not on the Teams page yet, please let us know and we can add you!