



# KABIR AHMAD

**Presentation**  
**Xaxor AI Bootcamp**

**29/05/2024**

# ADVANCED PRE PROCESSING TECHNIQUES

1

**SEMANTIC  
ROLE LABELING  
(SRL)**

2

**CONTEXTUAL  
EMBEDDING**

3

**NAMED ENTITY  
RECOGNITION  
(NER)**

4

**NATURAL  
LANGUAGE  
INFERENCE  
(NLI)**

# SEMANTIC ROLE LABELING (SRL)

The process of identifying the **predicate-argument structure of a sentence**, essentially determining who did what to whom, when, where, etc.

PREDICATE  
IDENTIFICATION

ARGUMENT  
IDENTIFICATION

ARGUMENT  
CLASSIFICATION

*SRL helps in understanding the deeper meaning of a sentence by identifying these relationships.*

# SEMANTIC ROLE LABELING (SRL)

LIBRARY

" ALENNLP "

SENTENCE

*"The chef cooked a delicious meal for the guests in the evening."*

RESULT

**Predicate:** cooked

**Arguments:**

[ARG0: The chef] [v: cooked]  
[ARG1: a delicious meal]  
[ARG2: for the guests]  
[ARGM-TMP: in the evening]

# CONTEXTUAL EMBEDDING

Contextual Embedding refers to **word embeddings** that capture the meaning of a word in the context of a sentence, as opposed to static embeddings like **Word2Vec** or **GloVe** which assign a single vector to a word regardless of its context.

**ELMO**

## Embeddings from Language Models

Generates embeddings for words considering their surrounding words. It uses bidirectional LSTM networks to create context-aware embeddings.

**BERT**

## Bidirectional Encoder Representations from Transformers

Uses a transformer architecture to generate embeddings that consider the context from both directions

# CONTEXTUAL EMBEDDING

LIBRARY

"TRANSFORMERS, AUTOTOKENIZER"

SENTENCE

*"The chef cooked a delicious meal for the guests in the evening."*

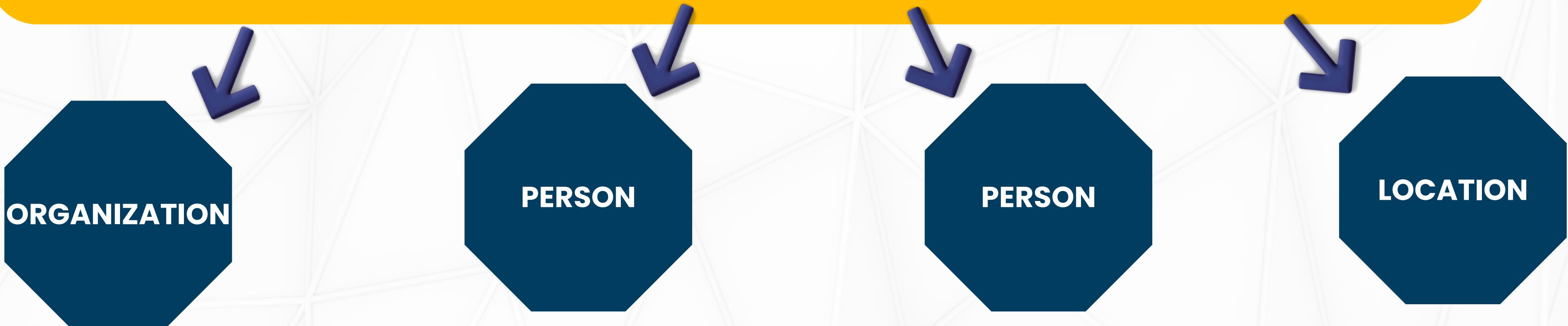
CONTEXTUAL  
EMBEDDINGS

```
tensor([[[ 0.1370, -0.1916, -0.4349, ..., -0.1848, 0.1336, -0.2597],  
        [-0.4665, -0.1544, -0.5044, ..., -0.2098, 0.4982, 0.0161],  
        [ 0.2236, 0.0223, 0.4061, ..., 0.3373, 0.2414, 0.1275],  
        ...,  
        [ 0.0503, -0.0612, 0.1524, ..., 0.0045, 0.3495, -0.0895],  
        [-0.3473, 0.0338, -0.1118, ..., 0.0542, -0.2061, -0.0311],  
        [-0.0578, 0.0362, -0.2267, ..., 0.1051, 0.0887, -0.0371]]])
```

# NAMED ENTITY RECOGNITION (NER)

The task of **identifying and classifying entities** in text into predefined categories such as names of persons, organizations, locations, dates, etc.

**" Google was founded by Larry Page and Sergey Brin in California "**



# NATURAL LANGUAGE INFERENCE (NLI)

Commonly known as **Recognizing Textual Entailment** (RTE), the task of determining the logical relationship between two sentences.

## ENTAILMENT

The truth of one sentence implies the truth of the other (e.g., "A dog is running" entails "An animal is running").

## CONTRADICTION

One sentence contradicts the other (e.g., "A dog is running" contradicts "No animals are running").

## NEUTRAL

The truth of one sentence does not affect the truth of the other (e.g., "A dog is running" and "It is raining" are unrelated).



# NATURAL LANGUAGE INFERENCE (NLI)

LIBRARY

"TRANSFORMERS, PIPELINE"

PREMISE

*"The chef cooked a delicious meal."*

HYPOTHESIS

*"The guests ate a delicious meal."*

RESULTS

Natural Language Inference Results:

Label: neutral, Score: 0.37

Label: contradiction, Score: 0.34

Label: entailment, Score: 0.29

# COMPARISON

Feature	NLTK	spaCy
Core NLP Tasks	Tokenization, POS tagging, parsing, NER	Tokenization, POS tagging, parsing, NER
Corpora and Lexical Resources	Extensive (e.g., WordNet, large corpora)	Limited
Pre-trained Models	Basic models	Advanced models (e.g., transformers)
Ease of Use	Moderate complexity	User-friendly, concise API
Performance	Slower, suitable for research and learning	Fast, optimized for production
Flexibility	Highly flexible, customizable	Less flexible, more opinionated
Integration	Standalone, less integration with ML libraries	Integrates with TensorFlow, PyTorch, etc.
Educational Resources	Extensive, ideal for teaching and research	Good documentation, less focus on teaching