NLP WORKSHOP SERIES
AN LLM INSIGHT

GPT

T5 (TTTTT)

BERT

TRANSFORMERS

The Path to Transformers

BOW	TF-IDF	Word2Vec	RNN	LSTM	Transformers
Counts words	Relative frequencies	Learns relationships	Remembers order	Remembers order	Remembers order
✓ Simple and fast	✓ Word importance	✓ Word similarity	✓ Relies on sentences	✓ Relies on sentences	✓ Relies on sentences
✗ Word meaning	✗ Word meaning	✓ Word meaning	✓ Context	✓ Context	✓ Context
✗ Word order	✗ Word order	✗ Context	✓ Short-term dependencies	✓ Short-term dependencies	✓ Short-term dependencies
		✗ Out Of Vocabulary (OOV)	✗ Forgets older words	✓ Remembers older words	✓ Remembers older words
				✗ Reads word by word (slow)	✓ Whole sentence at once

BOW

S1 = The rock band played an amazing live concert at the grand stadium.

S2 = She slipped on a rock near the river and saw a band of birds live in the trees of the grand forest.

X1 = [1 1 0 1 1 0 1 0 1 0 1 0 0 0 1 0 1 0 0 0 1 2 0]

X2 = [0 0 1 0 1 1 0 1 1 1 1 1 2 1 0 1 1 1 1 1 0 3 1]

cosine_similarity(X1, X2) = 0.50

The Path to Transformers

BOW	TF-IDF	Word2Vec	RNN	LSTM	Transformers
Counts words	Relative frequencies	Learns relationships	Remembers order	Remembers order	Remembers order
✓ Simple and fast	✓ Word importance	✓ Word similarity	✓ Relies on sentences	✓ Relies on sentences	✓ Relies on sentences
✗ Word meaning	✗ Word meaning	✓ Word meaning	✓ Context	✓ Context	✓ Context
✗ Word order	✗ Word order	✗ Context	✓ Short-term dependencies	✓ Short-term dependencies	✓ Short-term dependencies
		✗ Out Of Vocabulary (OOV)	✗ Forgets older words	✓ Remembers older words	✓ Remembers older words
				✗ Reads word by word (slow)	✓ Whole sentence at once

TF-IDF

S1 = The rock band played an amazing live concert at the grand stadium.

S2 = She slipped on a rock near the river and saw a band of birds live in the trees of the grand forest.

X1 = [0.31 0.31 0. 0.31 0.22 0. 0.31 0. 0.22 0. 0.22 0. 0. 0. 0.31 0. 0.22 0. 0. 0. 0.31 0.44 0.]

X2 = [0. 0. 0.21 0. 0.15 0.21 0. 0.21 0.15 0.21 0.15 0.21 0.43 0.21 0. 0.21 0.15 0.21 0.21 0.21 0. 0.45 0.21]

cosine_similarity(X1, X2) = 0.34

The Path to Transformers

BOW	TF-IDF	Word2Vec	RNN	LSTM	Transformers
Counts words	Relative frequencies	Learns relationships	Remembers order	Remembers order	Remembers order
✓ Simple and fast	✓ Word importance	✓ Word similarity	✓ Relies on sentences	✓ Relies on sentences	✓ Relies on sentences
✗ Word meaning	✗ Word meaning	✓ Word meaning	✓ Context	✓ Context	✓ Context
✗ Word order	✗ Word order	✗ Context	✓ Short-term dependencies	✓ Short-term dependencies	✓ Short-term dependencies
		✗ Out Of Vocabulary (OOV)	✗ Forgets older words	✓ Remembers older words	✓ Remembers older words
				✗ Reads word by word (slow)	✓ Whole sentence at once

Word2Vec

S1 = Word Vectors =

The	[-0.017,	0.007,	0.170]
Rock	[0.215,	0.299,	-0.167]
Band	[-0.125,	0.246,	-0.051]
Played	[... ,	...,	...]
An	[... ,	...,	...]
Amazing	[... ,	...,	...]
Live	[-0.060,	0.095,	0.033]
Concert	[... ,	...,	...]
At	[... ,	...,	...]
The	[-0.017,	0.007,	0.170]
Grand	[0.300,	-0.310,	-0.237]
stadium	[-0.267,	-0.314,	0.243]

V1 = [0.008, -0.018, 0.023]

S2= Word Vectors =

She	[-0.029,	0.184,	-0.091]
Slipped	[-0.291,	-0.146,	-0.001]
Rock	[0.215,	0.299,	-0.167]
Band	[0.215,	0.299,	-0.167]
Live	[-0.060,	0.095,	0.033]
In	[... ,	...,	...]
The	[-0.017,	0.007,	0.170]
Trees	[... ,	...,	...]
Of	[... ,	...,	...]
The	[-0.017,	0.007,	0.170]
Grand	[0.300,	-0.310,	-0.237]
forest	[-0.267,	-0.314,	0.243]

V2 = [0.005, -0.018, 0.004]

cosine_similarity(V1, V2) = 0.78

The Path to Transformers

BOW	TF-IDF	Word2Vec	RNN	LSTM	Transformers
Counts words	Relative frequencies	Learns relationships	Remembers order	Remembers order	Remembers order
✓ Simple and fast	✓ Word importance	✓ Word similarity	✓ Relies on sentences	✓ Relies on sentences	✓ Relies on sentences
✗ Word meaning	✗ Word meaning	✓ Word meaning	✓ Context	✓ Context	✓ Context
✗ Word order	✗ Word order	✗ Context	✓ Short-term dependencies	✓ Short-term dependencies	✓ Short-term dependencies
		✗ Out Of Vocabulary (OOV)	✗ Forgets older words	✓ Remembers older words	✓ Remembers older words
				✗ Reads word by word (slow)	✓ Whole sentence at once

RNN

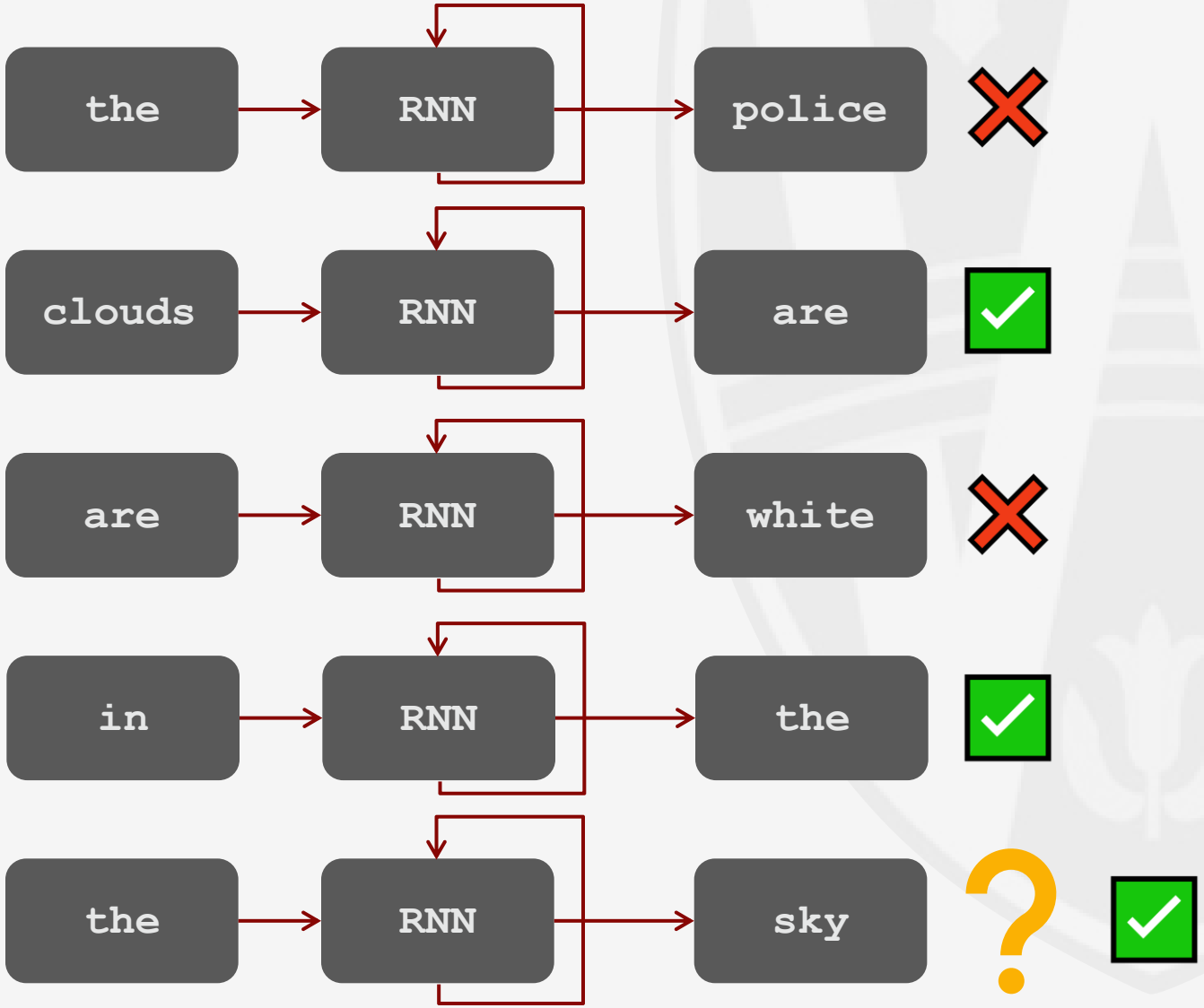
the ~~clouds are in the …~~

the clouds ~~are in the …~~

the clouds are ~~in the …~~

the clouds are in ~~the …~~

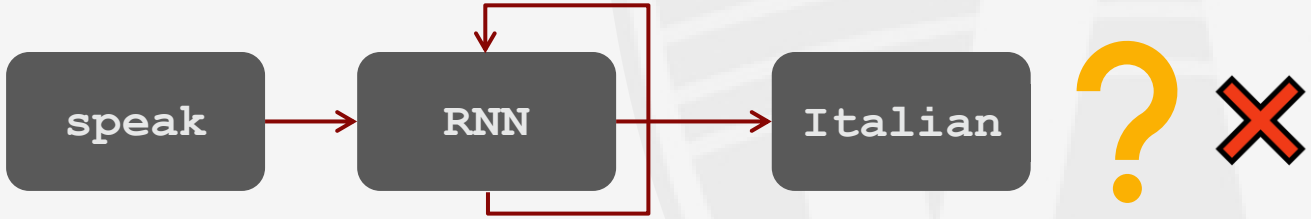
the clouds are in the …



RNN

I was born and raised in France where I spent 20 years of my life , and now it's hard for me to find a job in Italy because I speak … .

I was born and raised in France where I spent 20 years of my life , and now it's hard for me to find a job in Italy because I speak …



Vanishing Gradient Problem!

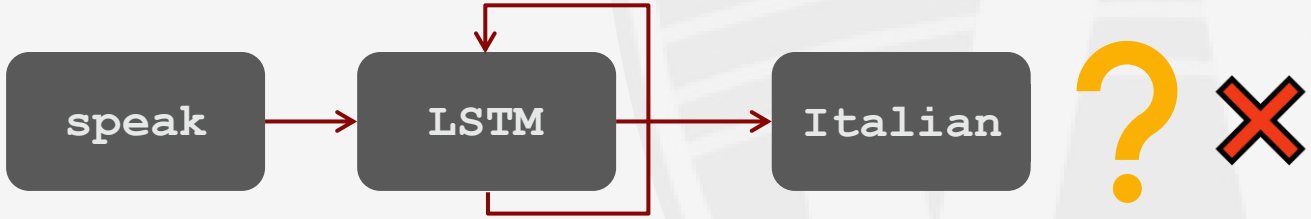
The Path to Transformers

BOW	TF-IDF	Word2Vec	RNN	LSTM	Transformers
Counts words	Relative frequencies	Learns relationships	Remembers order	Remembers order	Remembers order
✓ Simple and fast	✓ Word importance	✓ Word similarity	✓ Relies on sentences	✓ Relies on sentences	✓ Relies on sentences
✗ Word meaning	✗ Word meaning	✓ Word meaning	✓ Context	✓ Context	✓ Context
✗ Word order	✗ Word order	✗ Context	✓ Short-term dependencies	✓ Short-term dependencies	✓ Short-term dependencies
		✗ Out Of Vocabulary (OOV)	✗ Forgets older words	✓ Remembers older words	✓ Remembers older words
			✗ Reads word by word (slow)	✗ Reads word by word (slow)	✓ Whole sentence at once

LSTM

I was born and raised in France where I spent 20 years of my life , and now it's hard for me to find a job in Italy because I speak … .

I was born and raised in France where I spent 20 ~~years of my~~ life , and ~~now~~ it's hard ~~for me to~~ find a job in Italy because I speak …



LSTMs were pioneering in almost all predicting and generating tasks since the introduction of Transformers!

The Path to Transformers

BOW	TF-IDF	Word2Vec	RNN	LSTM	Transformers
Counts words	Relative frequencies	Learns relationships	Remembers order	Remembers order	Remembers order
✓ Simple and fast	✓ Word importance	✓ Word similarity	✓ Relies on sentences	✓ Relies on sentences	✓ Relies on sentences
✗ Word meaning	✗ Word meaning	✓ Word meaning	✓ Context	✓ Context	✓ Context
✗ Word order	✗ Word order	✗ Context	✓ Short-term dependencies	✓ Short-term dependencies	✓ Short-term dependencies
		✗ Out Of Vocabulary (OOV)	✗ Forgets older words	✓ Remembers older words	✓ Remembers older words
				✗ Reads word by word (slow)	✓ Whole sentence at once

Transformers

Main problems with RNNs and LSTMs:

- They process the input data sequentially (word by word)
- They become quite ineffective when elements are distant from one another

Transformers address these problems by:

- Pay attention to specific words, no matter how distant they are
- Boost the performance speed by process the input data at once with parallel computation

Transformers

1. **Tokenization & Embedding** → Convert words to vectors (word embedding + positional embedding).
2. **Self-Attention** → Determine word relationships.
3. **Multi-Head Attention** → Capture multiple perspectives (action, time, location, ...).
4. **Feed-Forward Processing** → Refine representations for different tasks and applications.
5. **Output Generation** → Use the processed context for predictions.

Transformers - Tokenization & Embedding

"The cat sat on the mat."

Word Embeddings:

```
[0.2, 0.5, ...],  # The
[0.8, 1.2, ...],  # cat
[0.3, 0.9, ...],  # sat
[0.4, 0.7, ...],  # on
[0.2, 0.5, ...],  # the
[0.6, 1.0, ...],  # mat
```

Positional Embeddings:

```
[1],  # The
[2],  # cat
[3],  # sat
[4],  # on
[5],  # the
[6],  # mat
```

Transformers - Attention

"The cat sat on the mat."

Attention Mechanism:

```
Attention("cat" → "sat") = 0.8   # "cat" attends to "sat" (important)
Attention("cat" → "mat") = 0.4   # "cat" attends to "mat" (less important)
Attention("cat" → "on")  = 0.1   # "cat" does not focus on "on"
```

Transformers - Multi-Head Attention

"The cat sat on the mat."

Multi-Head Attention Mechanism:

"cat" → "sat" (**action** link)

"cat" → "mat" (**location** link)

Transformers - Feed-Forward Network

"The cat sat on the

mat."

refining the contextual understanding for a specific task

Translation:

- Input: The cat sat on the mat
- Output: Le chat s'est assis sur le tapis

Next word prediction:

- Input: The cat sat on the mat
- Output: ,drinking water