

# Машинное обучение

Лекция 12  
Modern methods of NLP

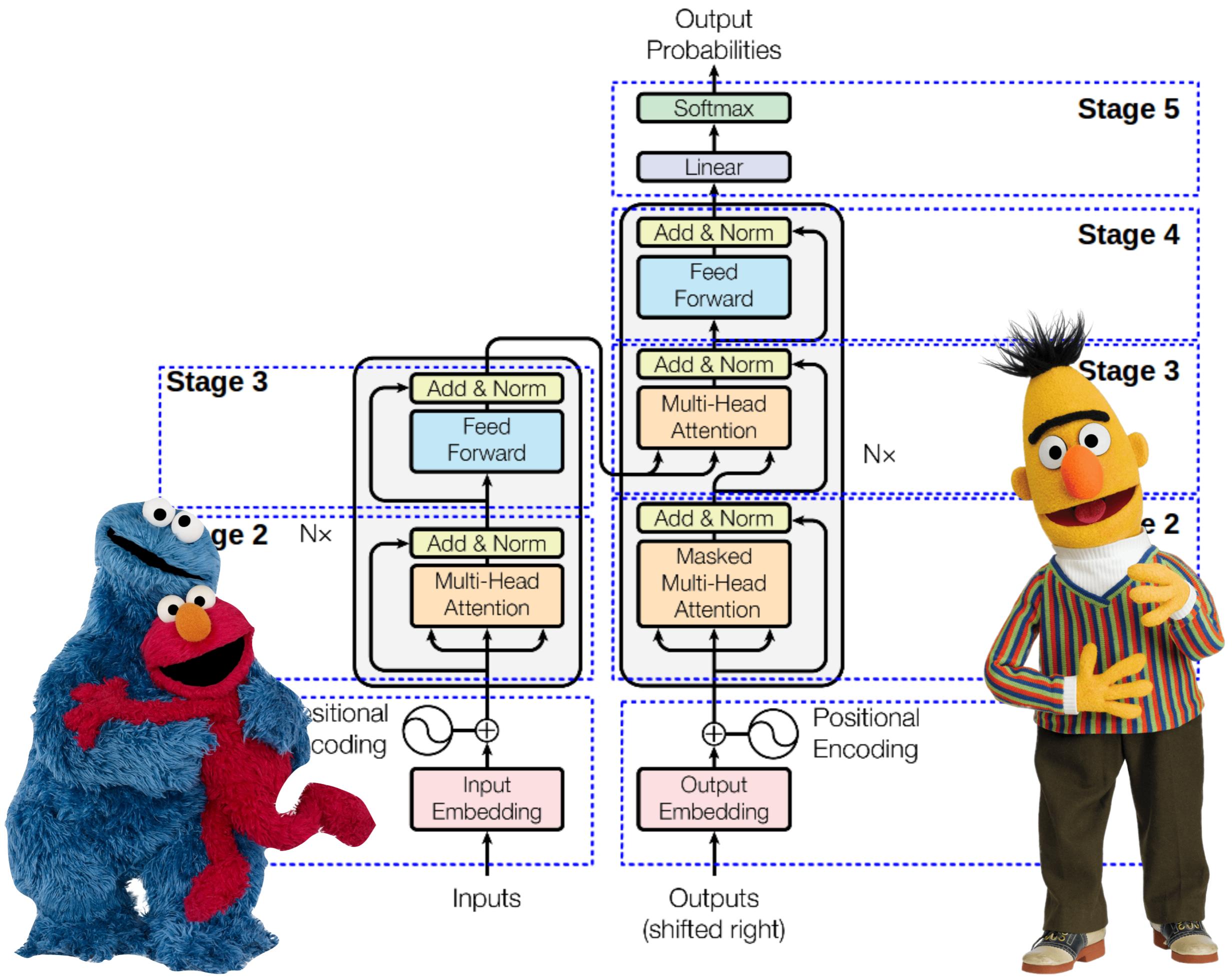
*Власов Кирилл Вячеславович*



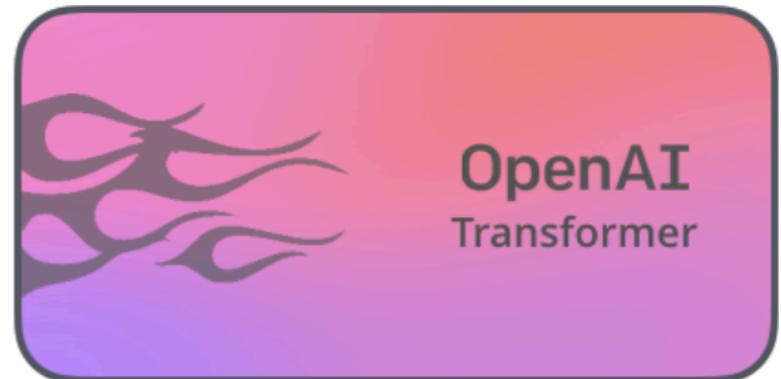
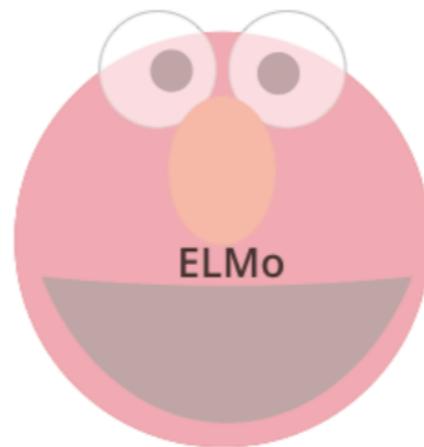
2019

# Self-Attention

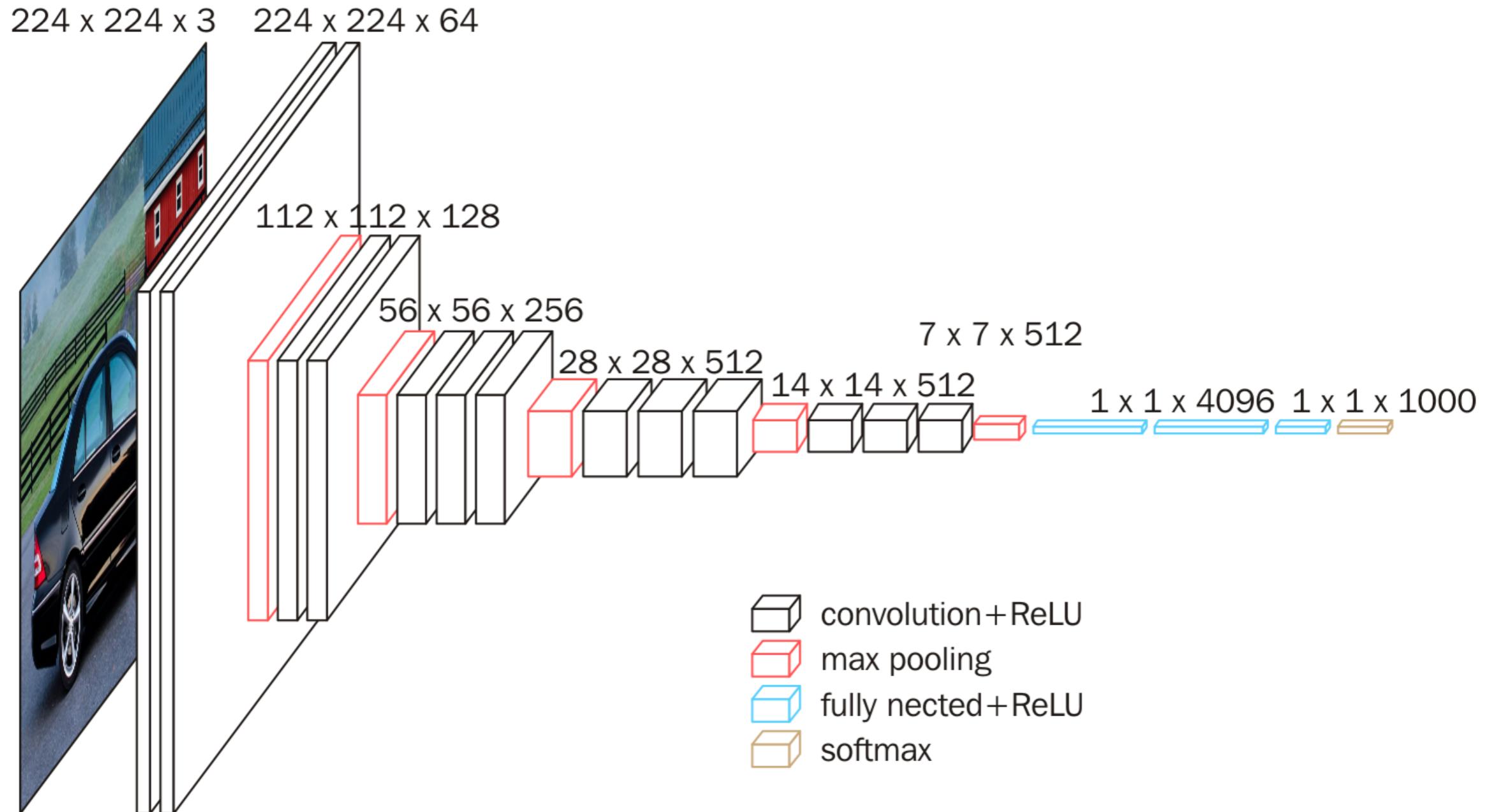




# План лекции

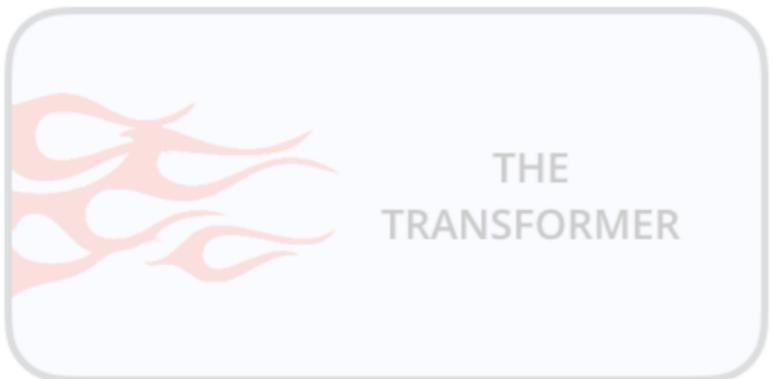


# Recap: Applying transfer learning in CV



# **Transfer learning in NLP?**

# План лекции



# Universal Language Model



Architecture that ULMFiT uses for its language modeling task is an AWD-LSTM. The name is an abbreviation of ASGD Weight-Dropped LSTM.

**YOU GET DROPOUT, YOU GET DROPOUT**



**EVERYBODY GETS DROPOUT**

# Universal Language Model

ENCODER DROPOUT				
	before dropout			after dropout
	token	d1	d2	...
	I	0.399	0.75379	0.62616
	love	0.88533	0.29449	0.15856
	cats	0.48927	0.04071	0.21427
	dogs	0.72918	0.86882	0.77136

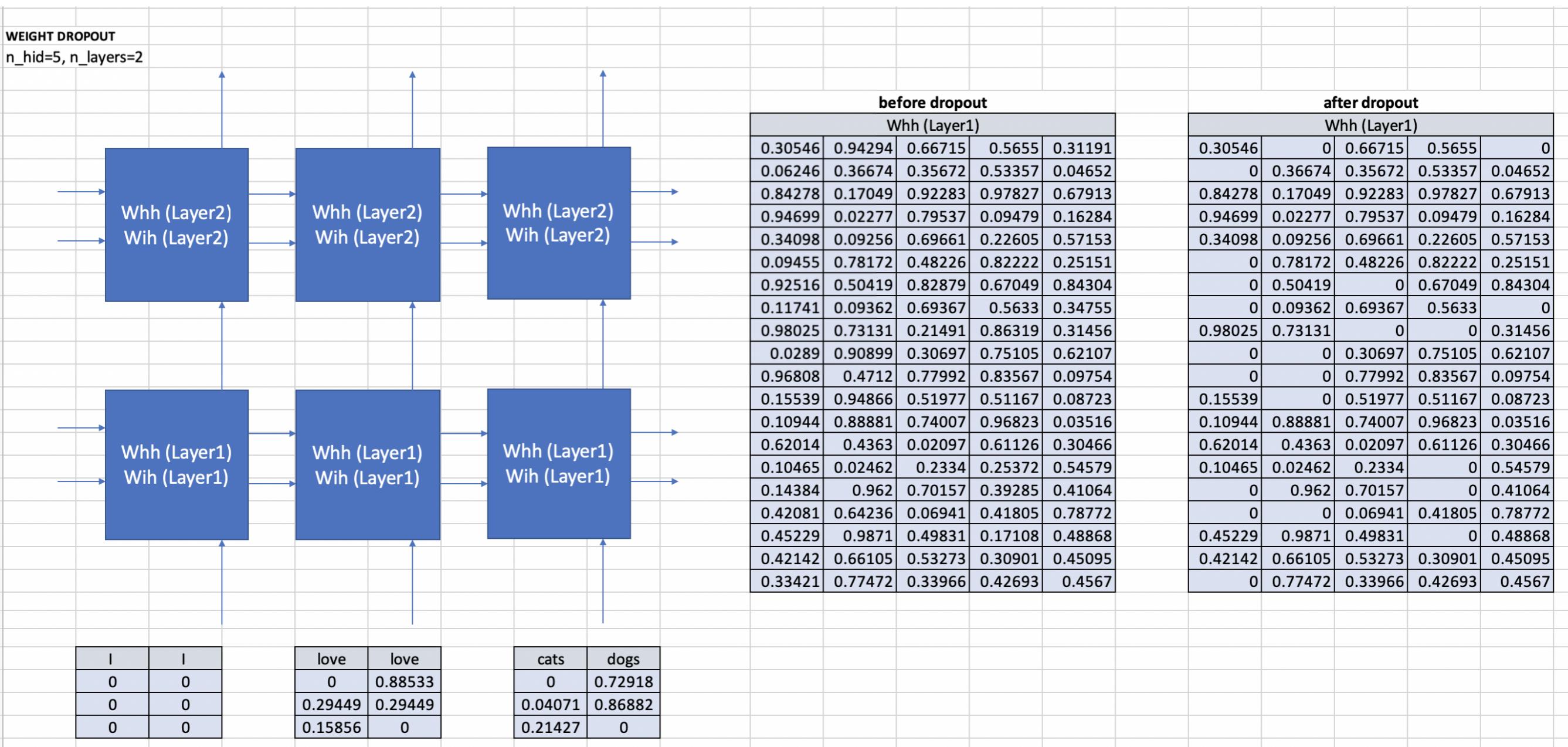
DECODER DROPOUT				
	before dropout			after dropout
	token	d1	d2	...
	I	0	0	0
	love	0.88533	0.29449	0.15856
	cats	0.48927	0.04071	0.21427
	dogs	0.72918	0.86882	0.77136

# Universal Language Model

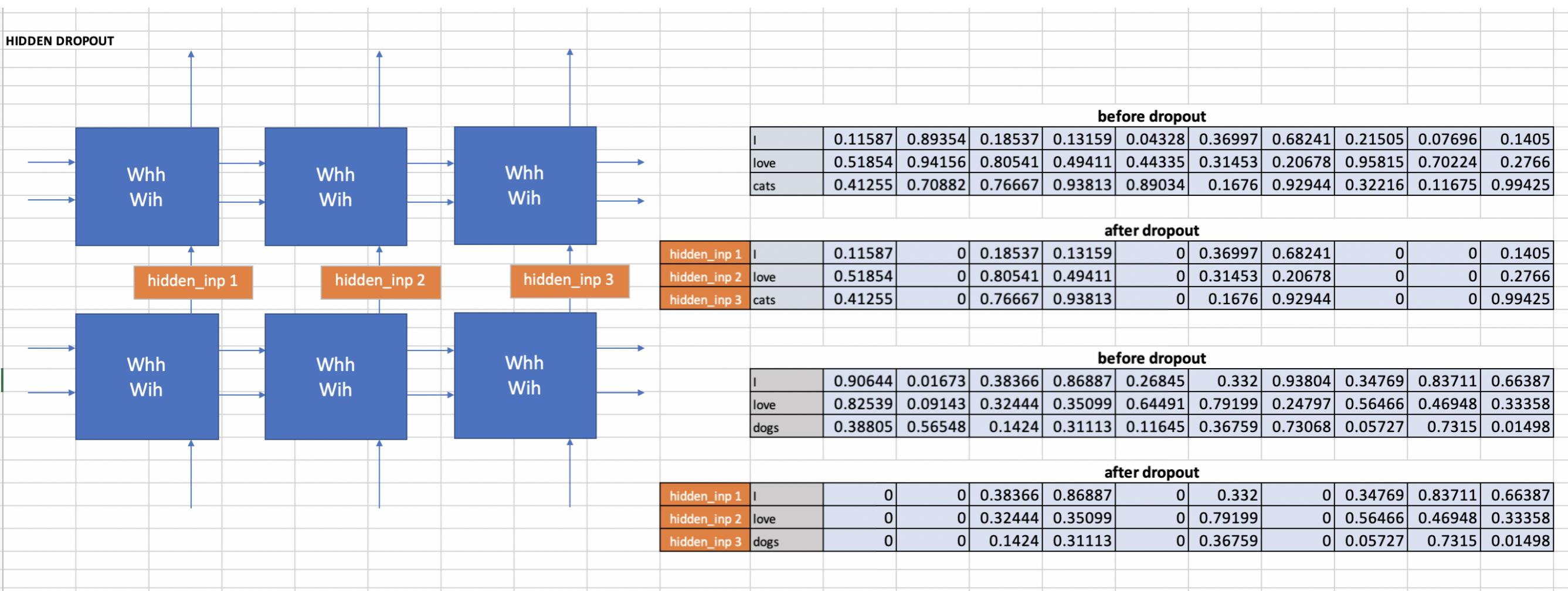
ENCODER DROPOUT				
	before dropout			after dropout
token	d1	d2	...	token
I	0.399	0.75379	0.62616	I
love	0.88533	0.29449	0.15856	love
cats	0.48927	0.04071	0.21427	cats
dogs	0.72918	0.86882	0.77136	dogs

INPUT DROPOUT				
batch: [I love cats, I love dogs]				
	before dropout			after dropout
I	0	0	0	I
love	0.88533	0.29449	0.15856	love
cats	0.48927	0.04071	0.21427	cats
	before dropout			after dropout
I	0	0	0	I
love	0.88533	0.29449	0.15856	love
dogs	0.72918	0.86882	0.77136	dogs

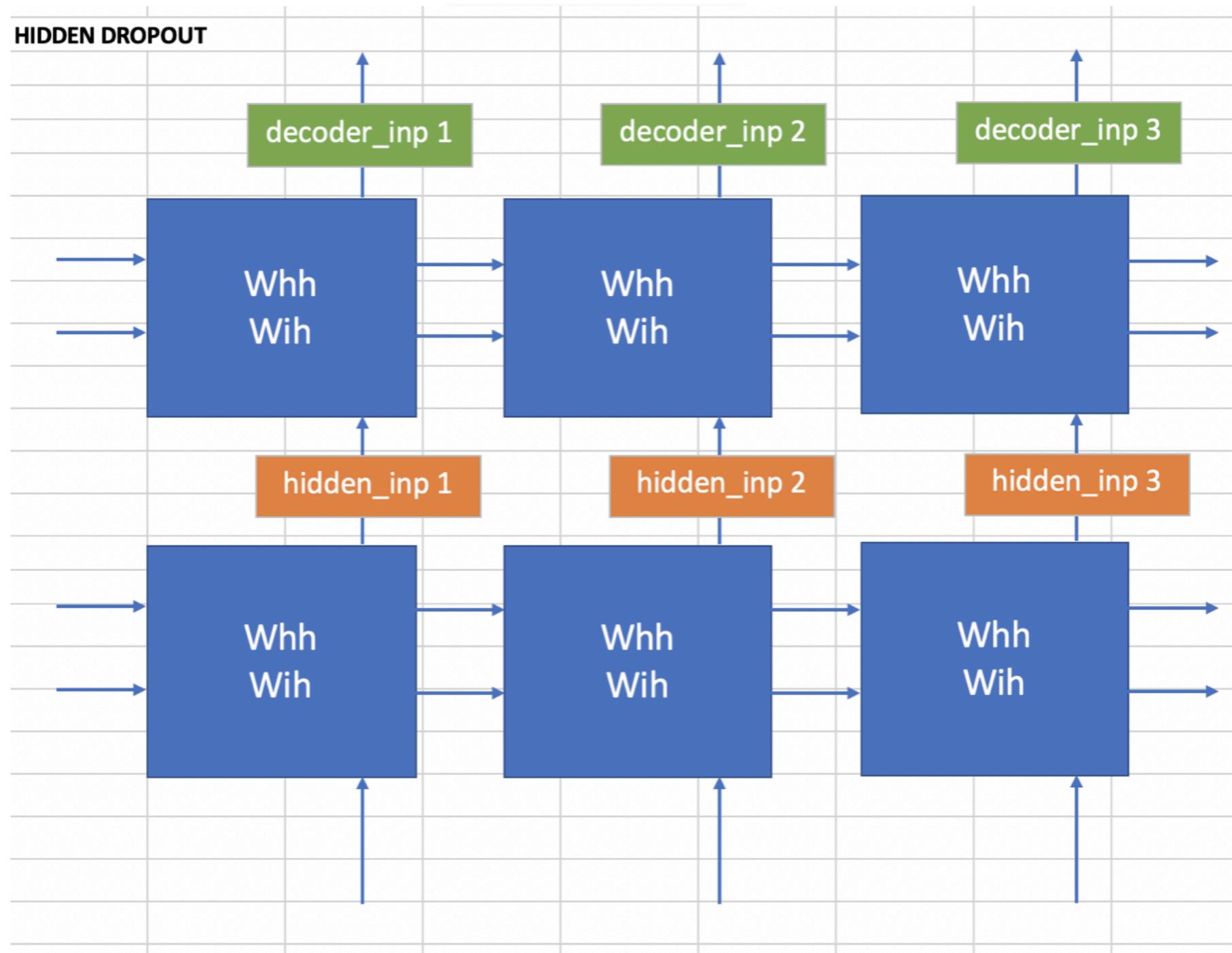
# Universal Language Model



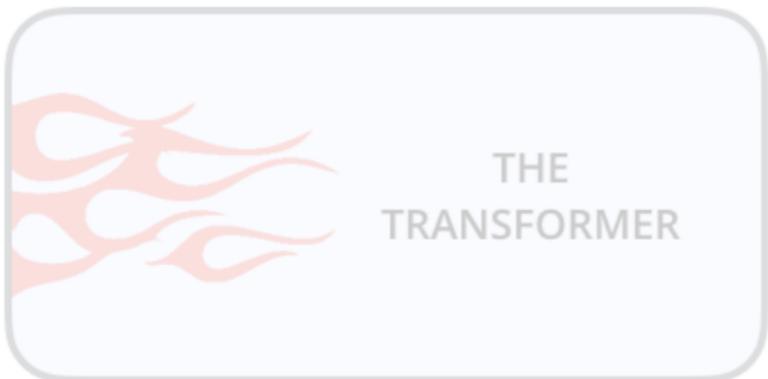
# Universal Language Model



# Universal Language Model



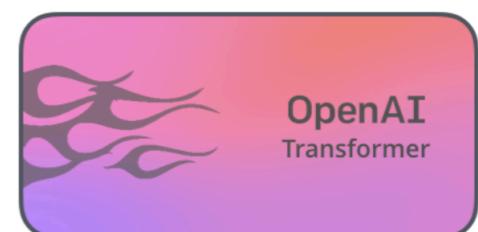
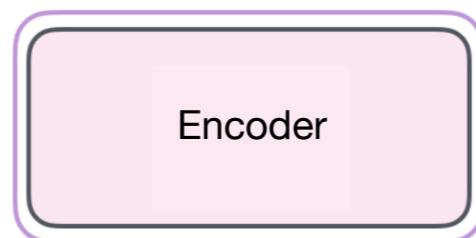
# План лекции



# OpenAI Transformer

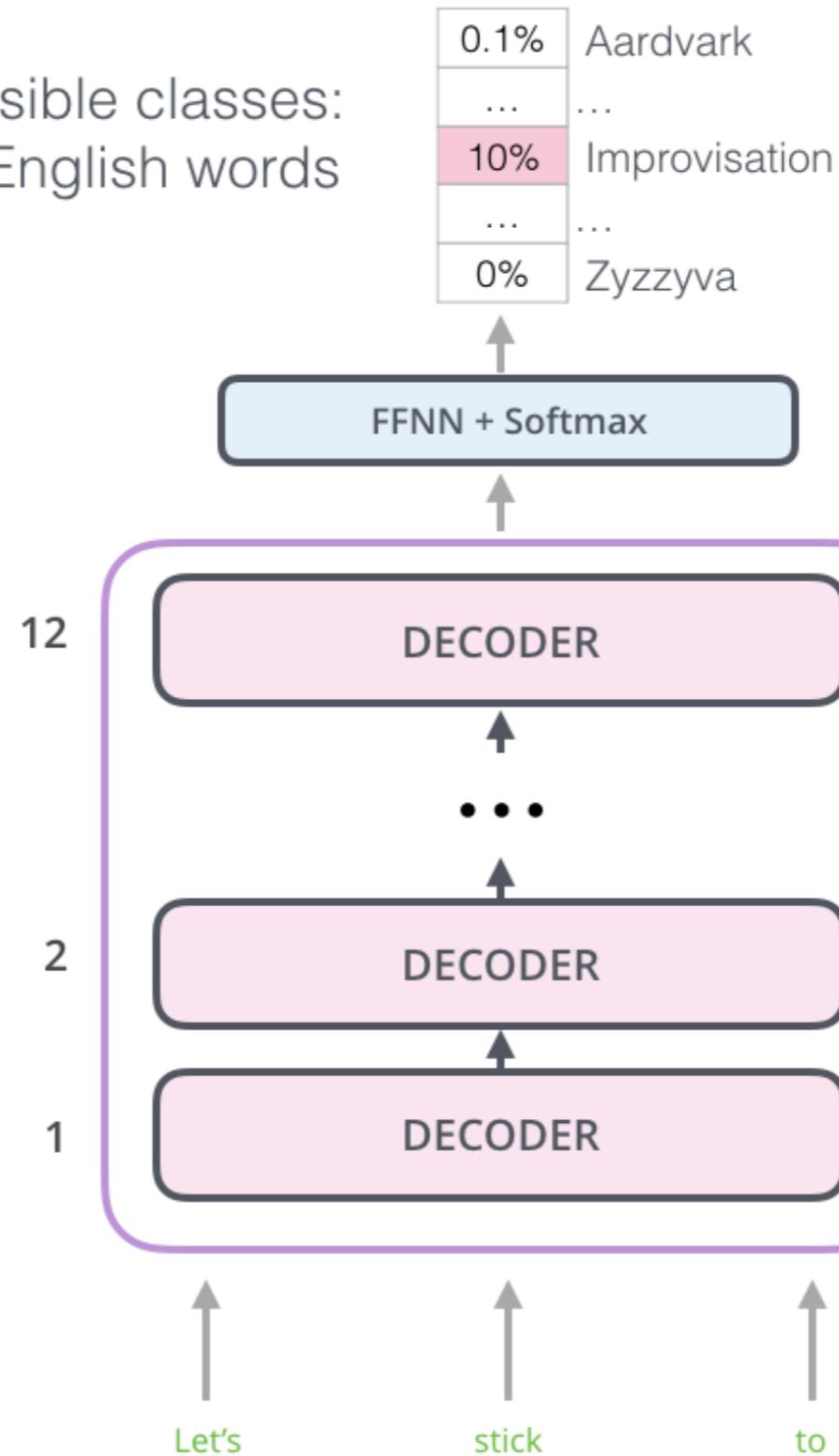


Pre-training a Transformer Decoder for  
Language Modeling

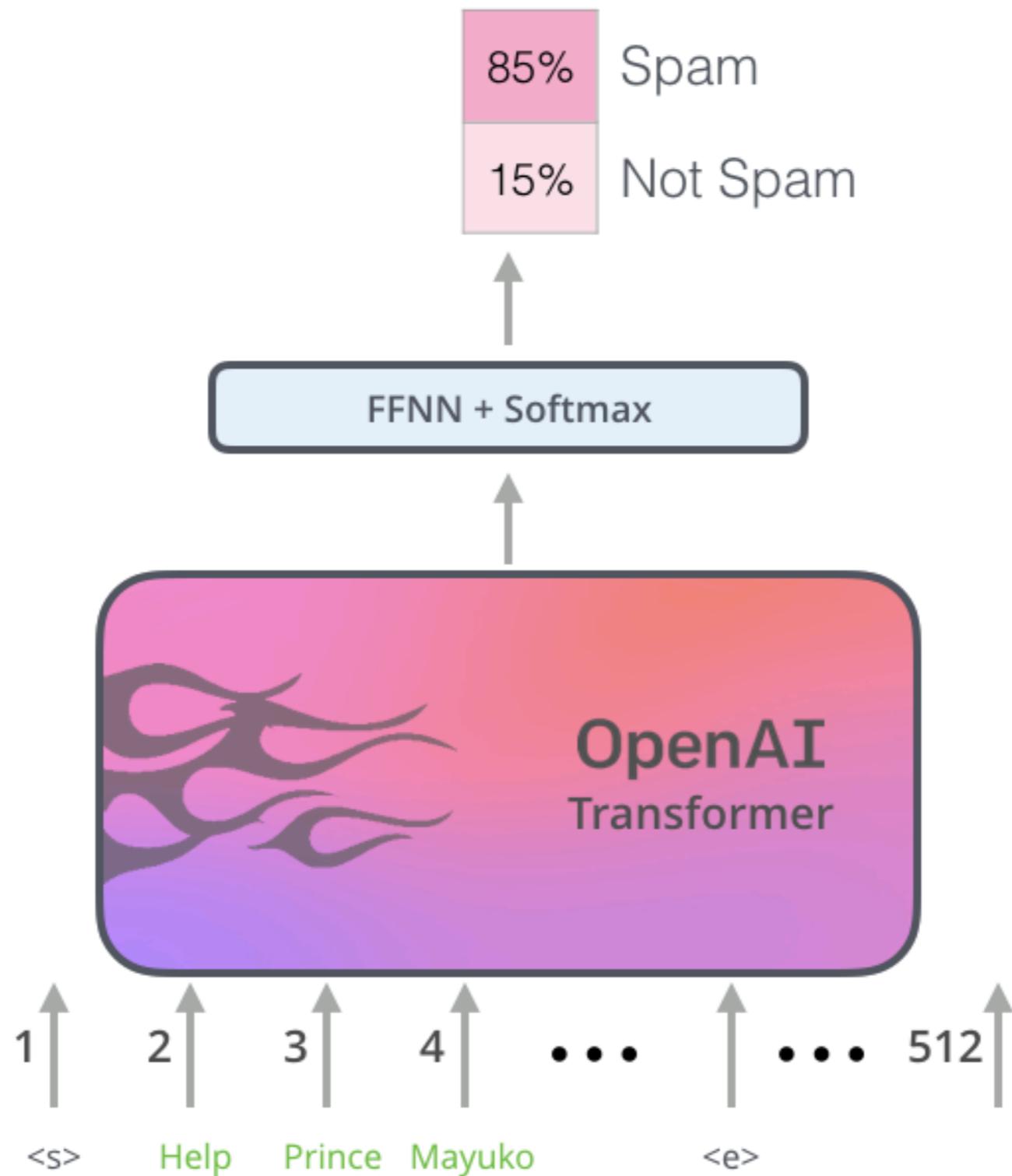


# OpenAI Transformer

Possible classes:  
All English words

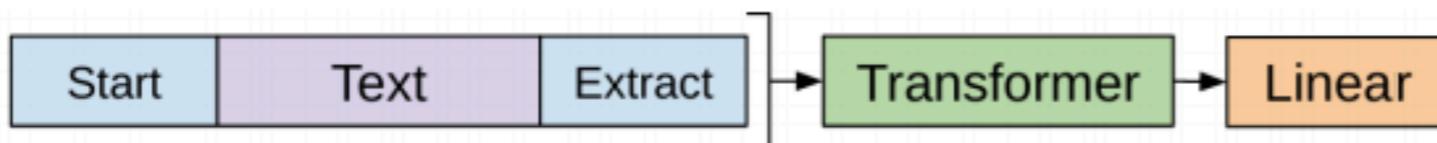


# OpenAI Transformer

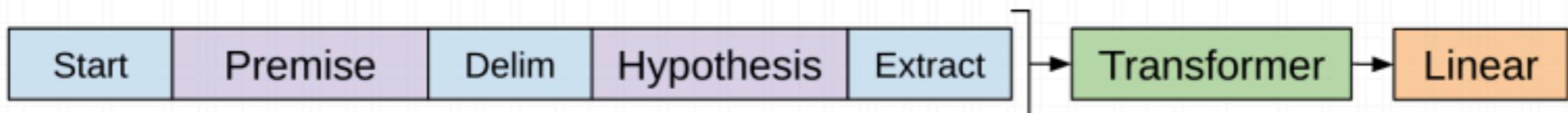


# OpenAI Transformer

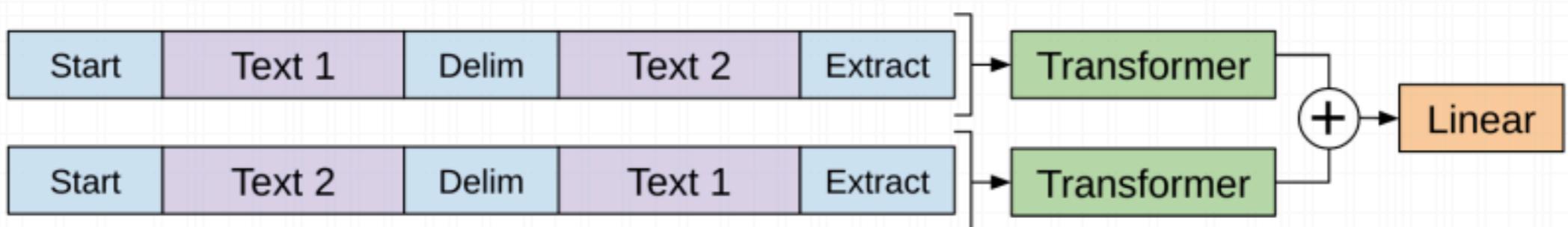
Classification



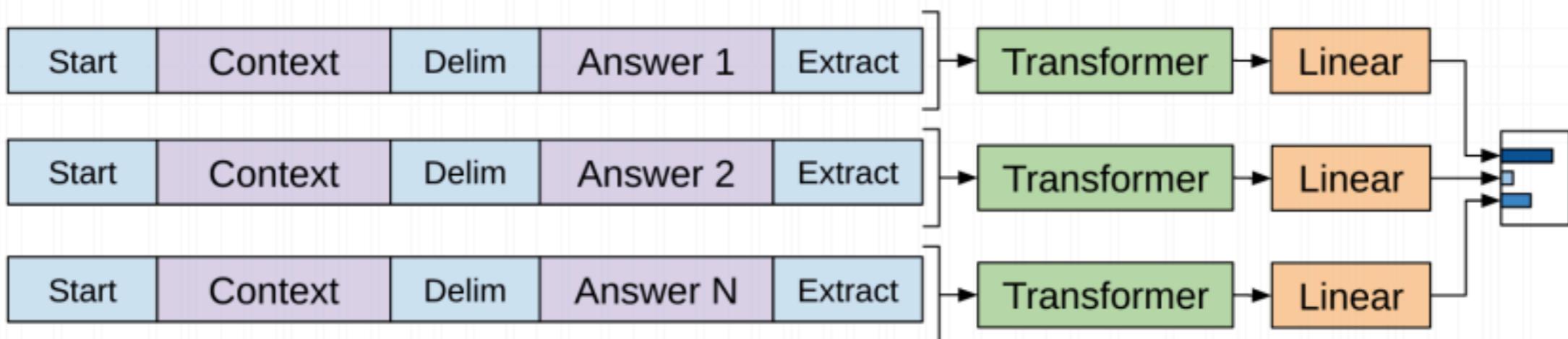
Entailment



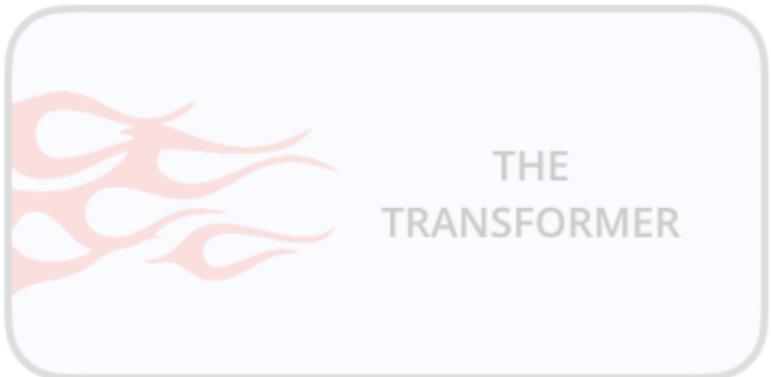
Similarity



Multiple Choice



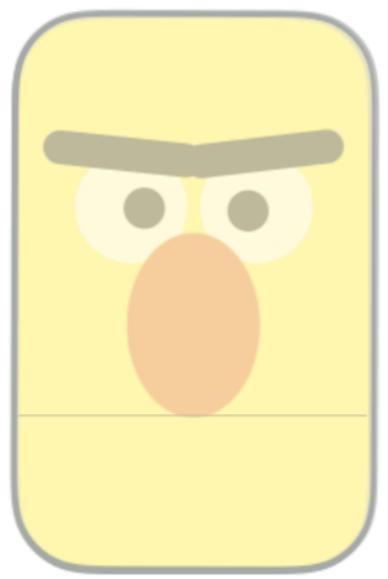
# План лекции



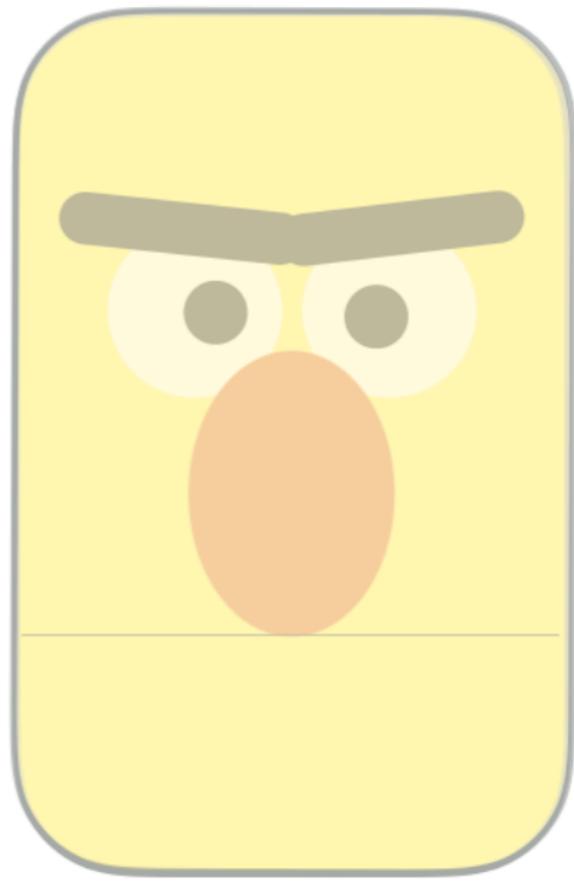
# BERT



# BERT

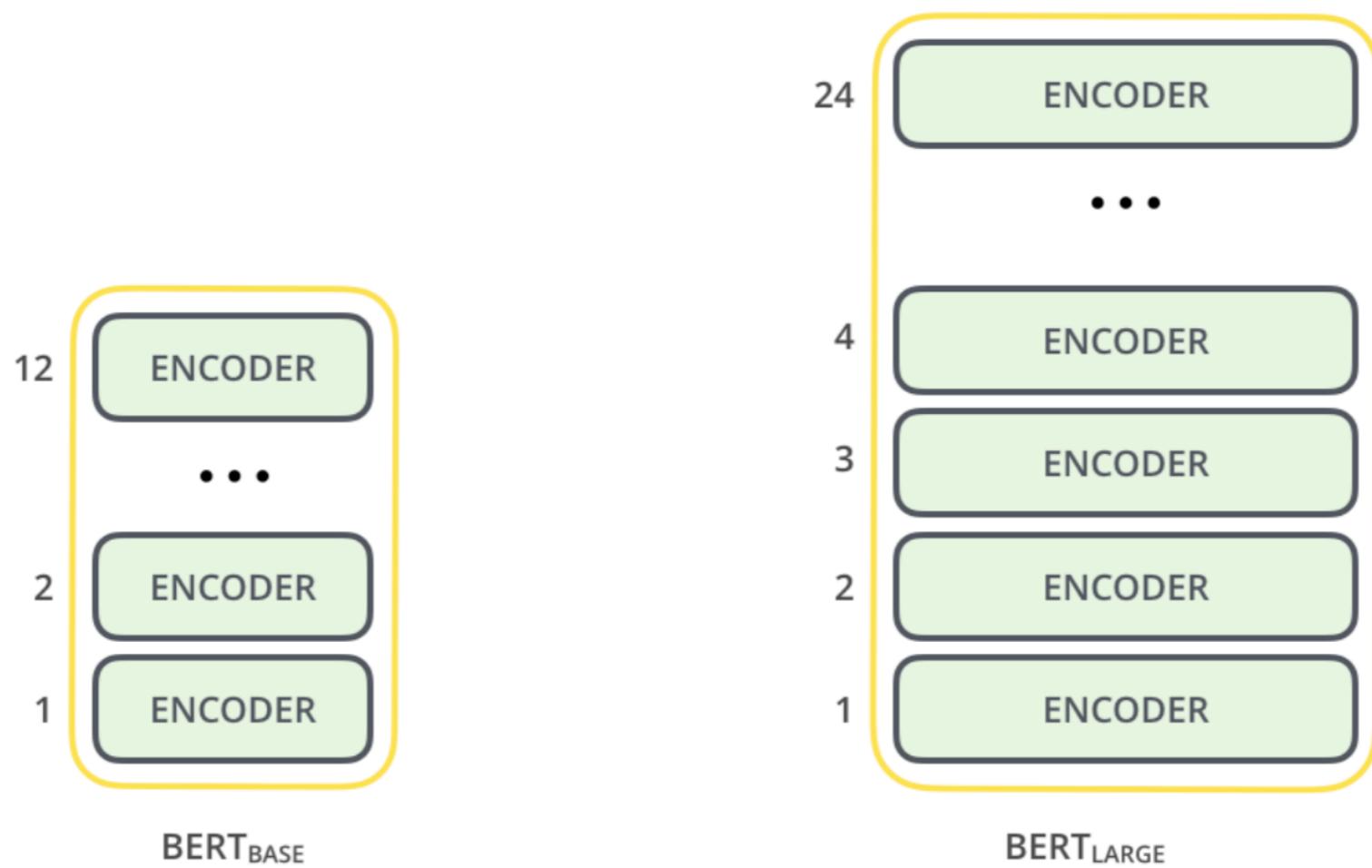


BERT<sub>BASE</sub>

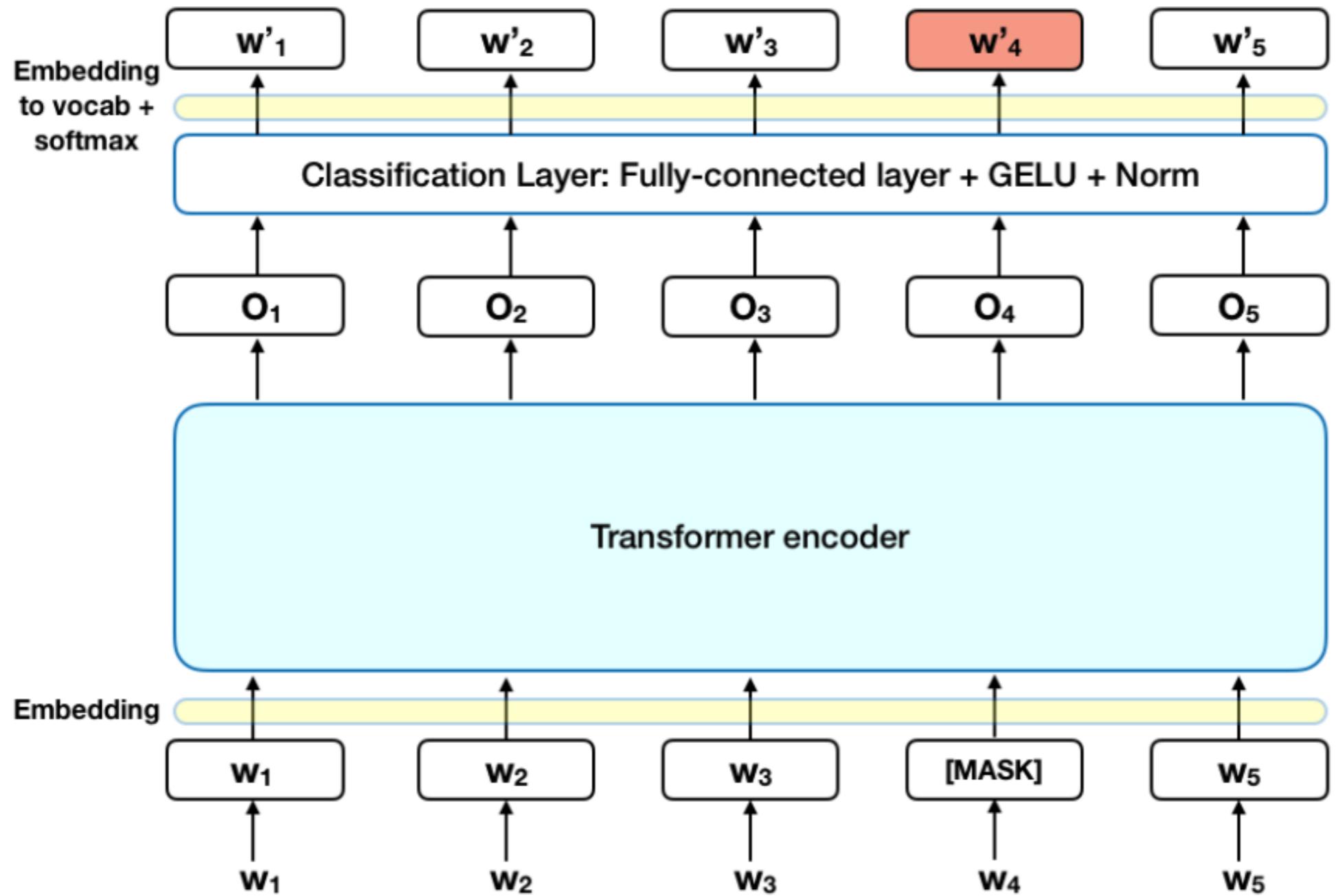


BERT<sub>LARGE</sub>

# BERT

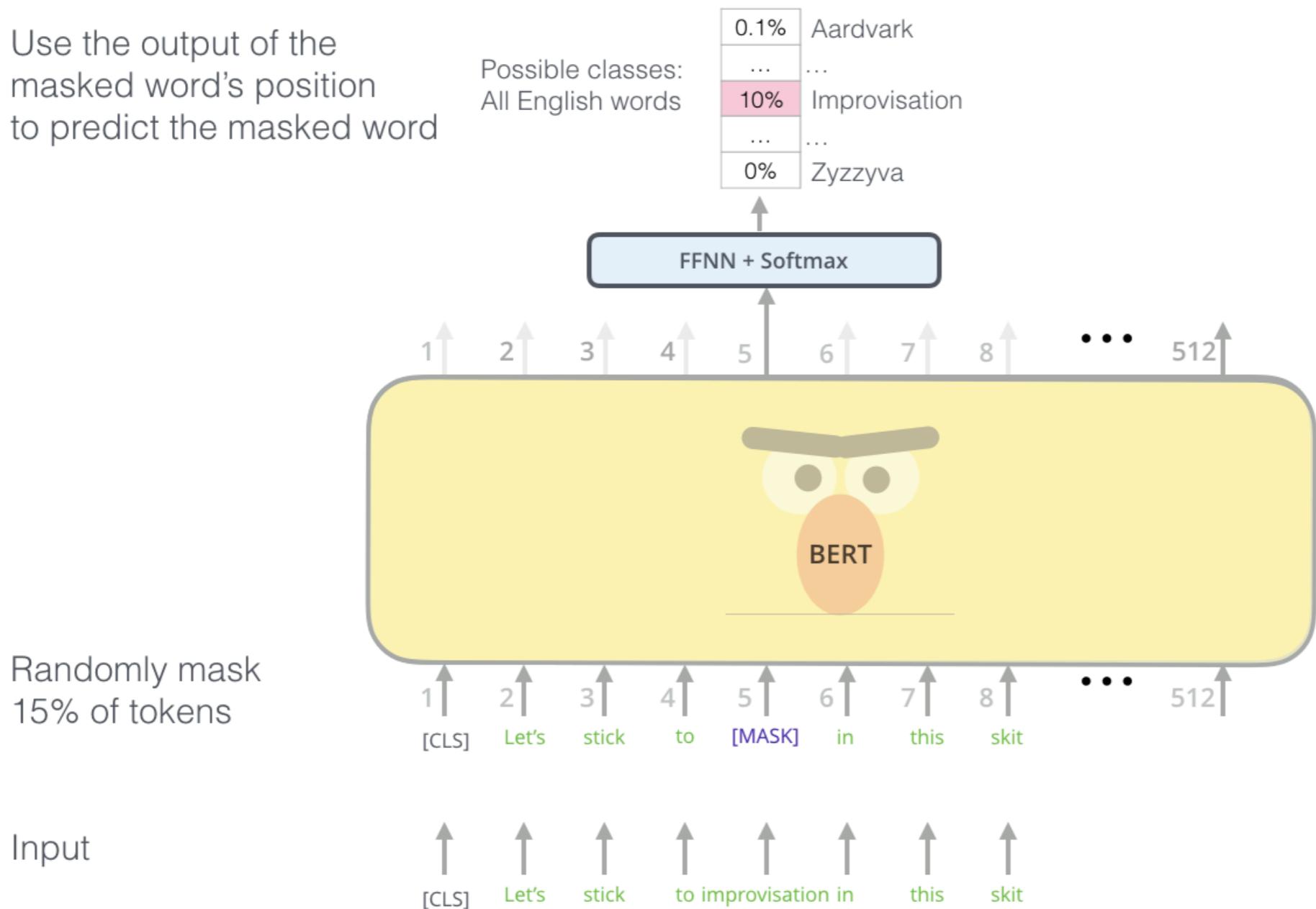


# BERT



# BERT

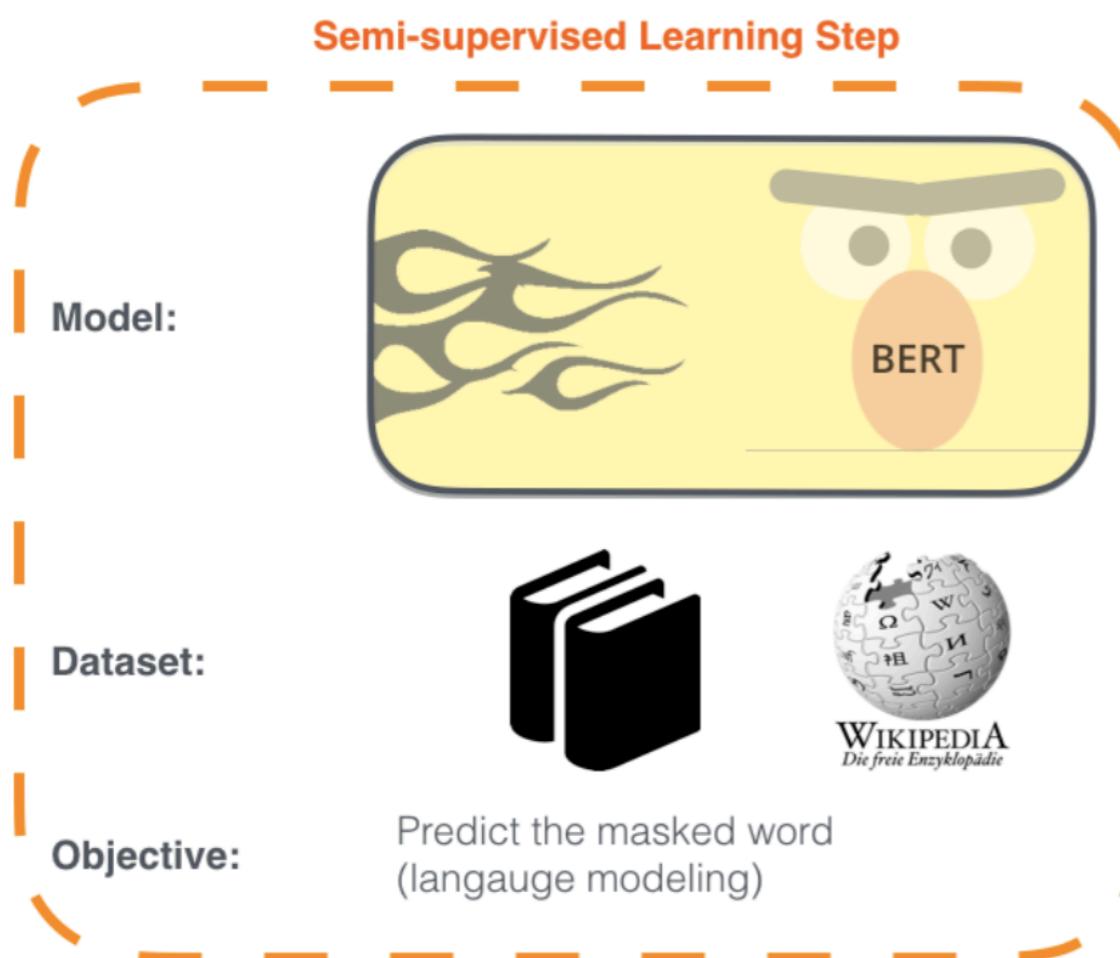
Use the output of the masked word's position to predict the masked word



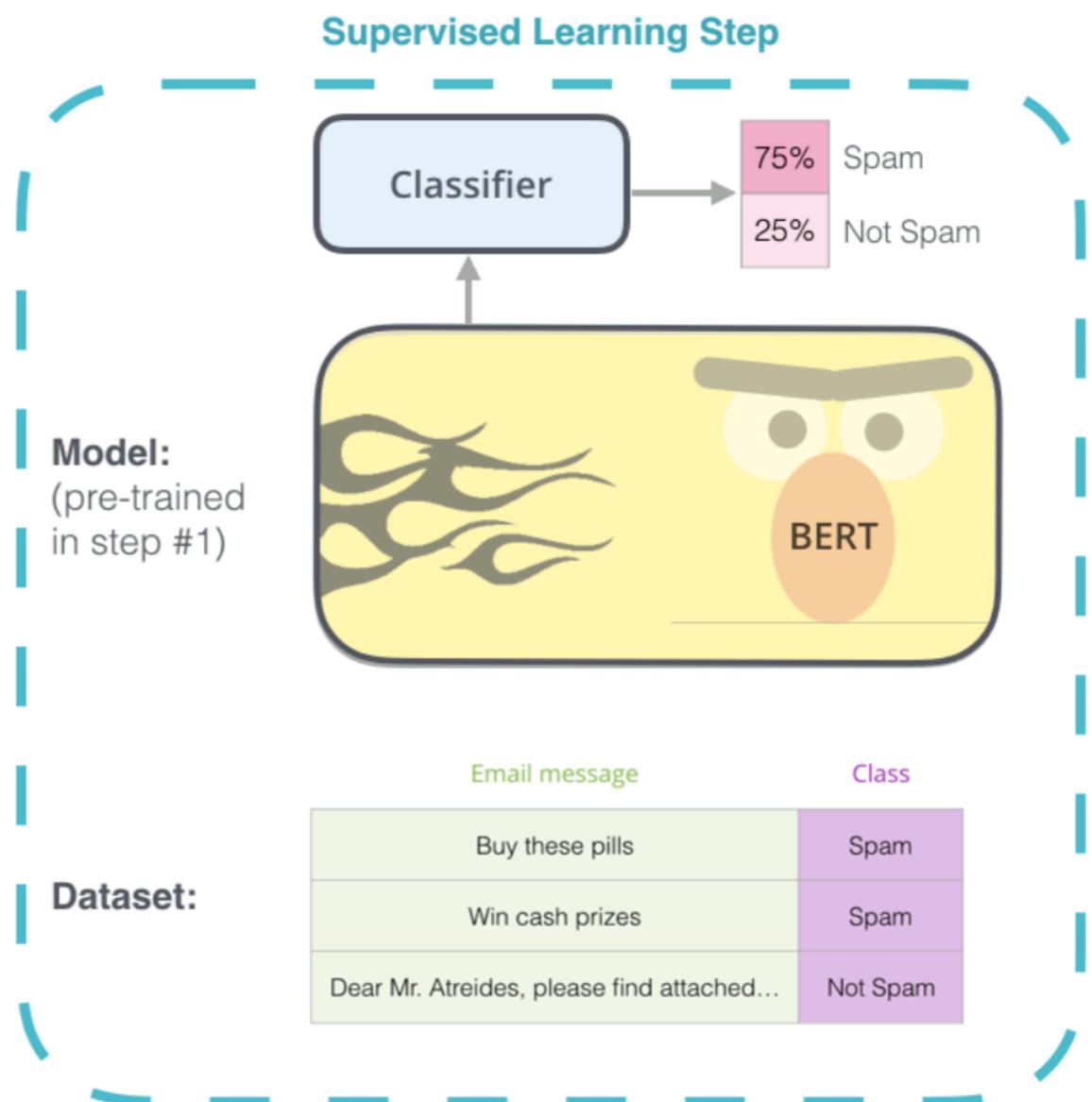
# BERT

1 - **Semi-supervised** training on large amounts of text (books, wikipedia..etc).

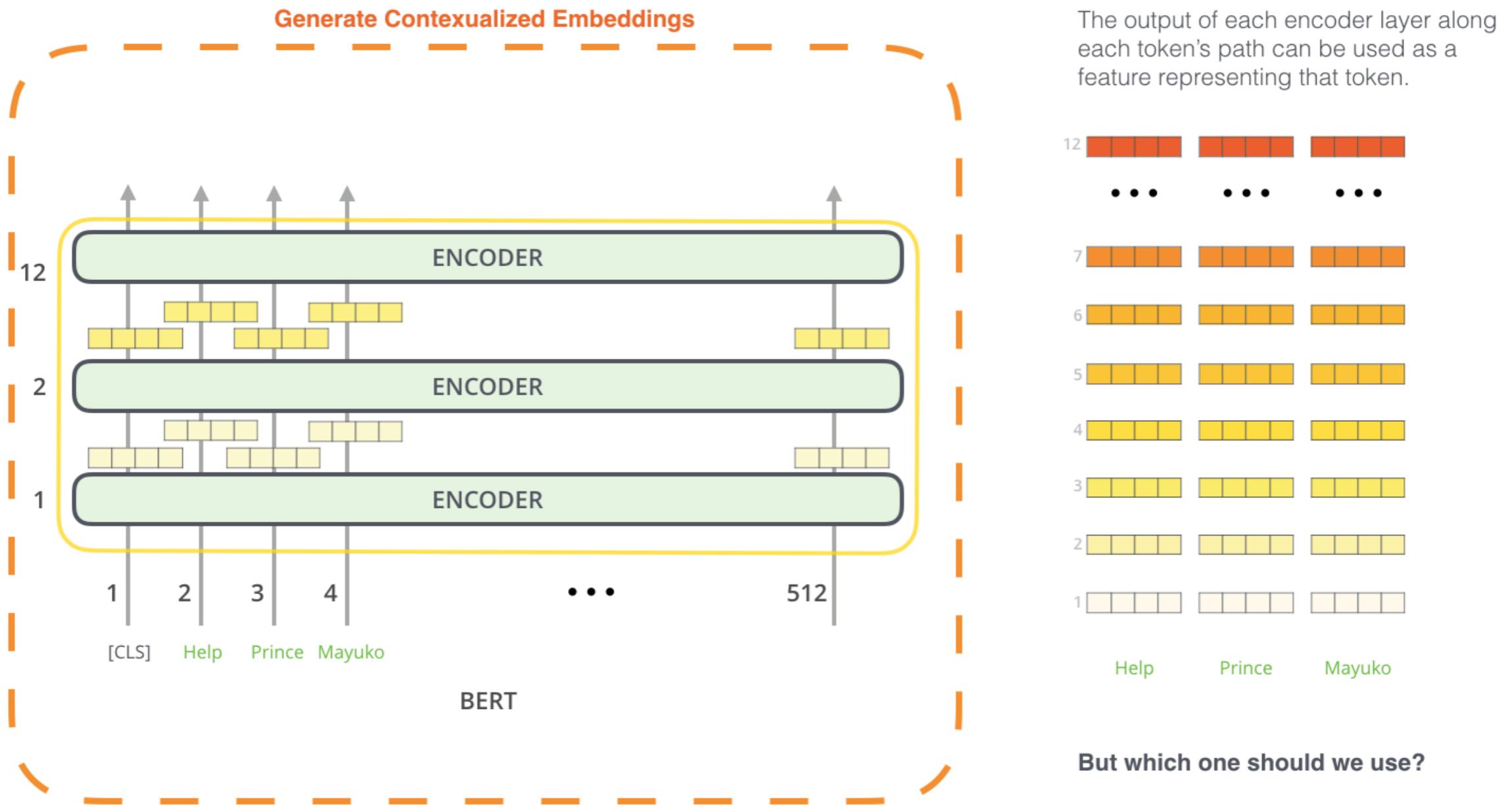
The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.



2 - **Supervised** training on a specific task with a labeled dataset.



# BERT for feature extraction



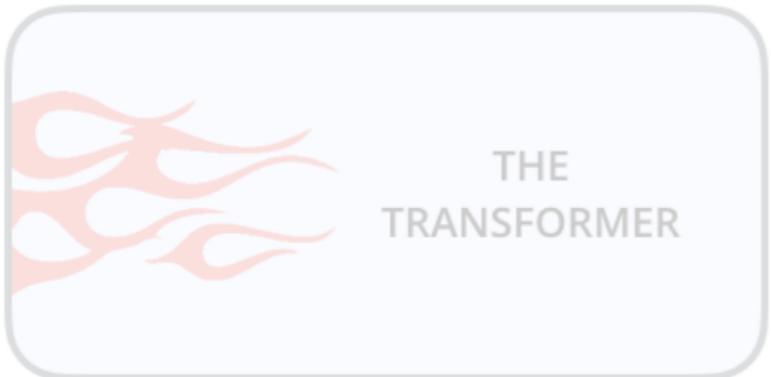
# BERT for feature extraction

What is the best contextualized embedding for “**Help**” in that context?

For named-entity recognition task CoNLL-2003 NER

		Dev F1 Score
12		91.0
• • •		
7		94.9
6		
5		
4		
3		95.5
2		
1		95.6
Help		
		95.9
		96.1

# План лекции



# ELMo



1. Expedited Labour Market Opinion
2. Electric Light Machine Organization
3. Enough Let's Move On

Special thx for @Anastasia Yanina:

[https://github.com/ml-mipt/ml-mipt/blob/advanced/week05\\_BERT\\_and\\_LDA/Lecture\\_BERT\\_DIHT.pdf](https://github.com/ml-mipt/ml-mipt/blob/advanced/week05_BERT_and_LDA/Lecture_BERT_DIHT.pdf)

# ELMo



1. Expedited Labour Market Opinion
  2. Electric Light Machine Organization
  3. Enough Let's Move On
- 4. Embeddings from Language Models**

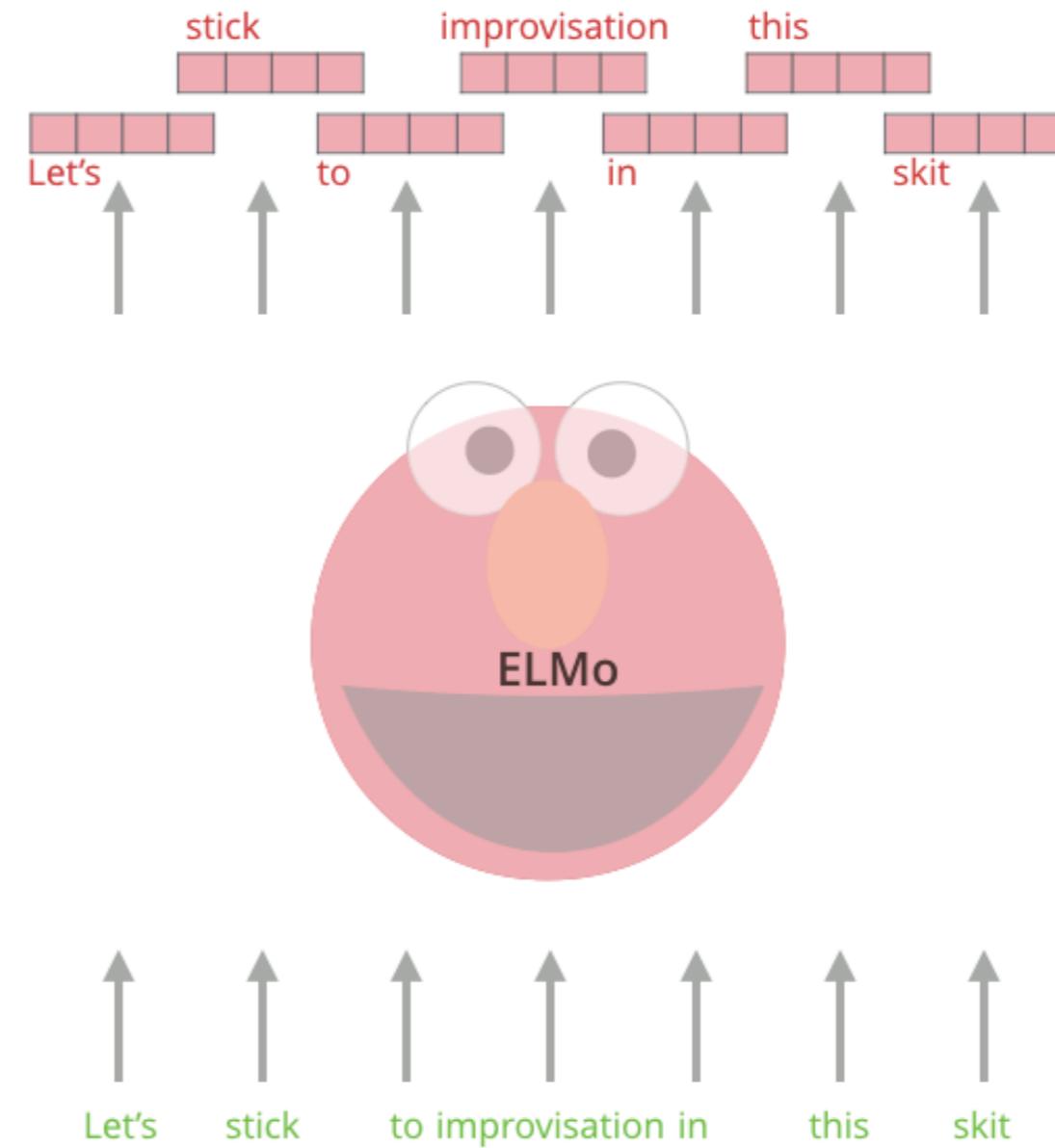
Special thx for @Anastasia Yanina:

[https://github.com/ml-mipt/ml-mipt/blob/advanced/week05\\_BERT\\_and\\_LDA/Lecture\\_BERT\\_DIHT.pdf](https://github.com/ml-mipt/ml-mipt/blob/advanced/week05_BERT_and_LDA/Lecture_BERT_DIHT.pdf)

# ELMo

ELMo  
Embeddings

Words to embed





Hey ELMo, what's the embedding  
of the word "stick"?

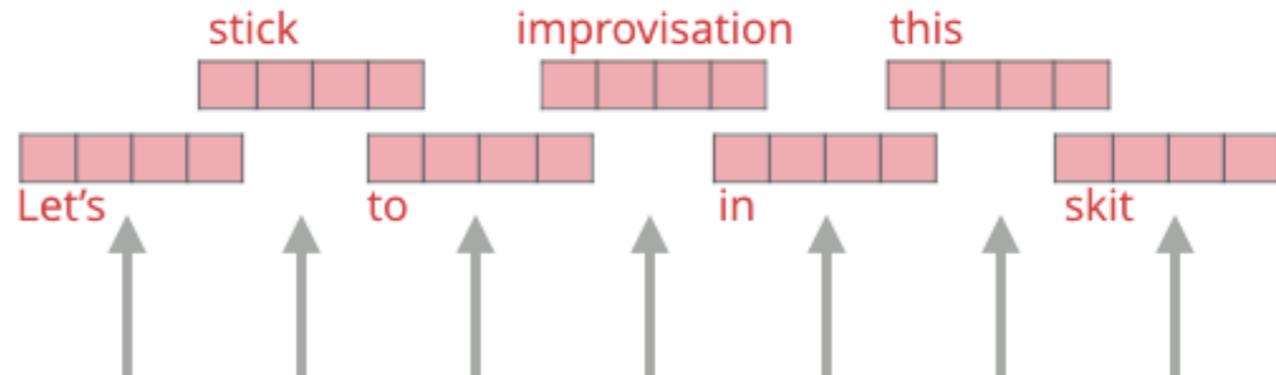
There are multiple possible  
embeddings! Use it in a sentence.

Oh, okay. Here:  
"Let's stick to improvisation in this  
skit"

Oh in that case, the embedding is:  
-0.02, -0.16, 0.12, -0.1 ....etc

# ELMo

ELMo  
Embeddings



Words to embed



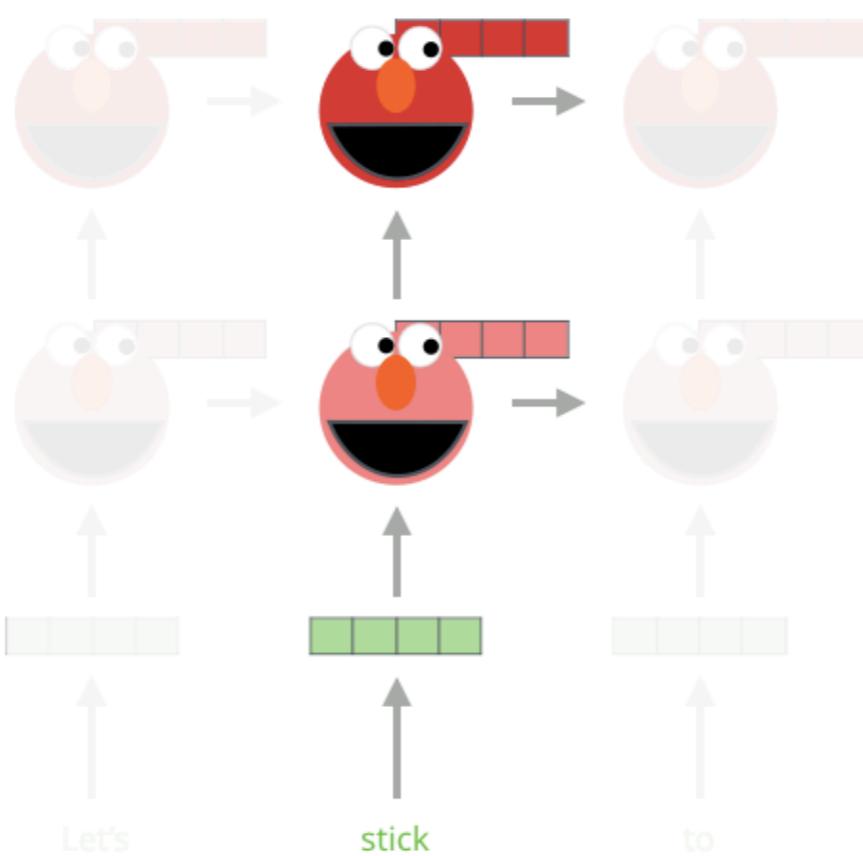
# ELMo

Embedding of “stick” in “Let’s stick to” - Step #2

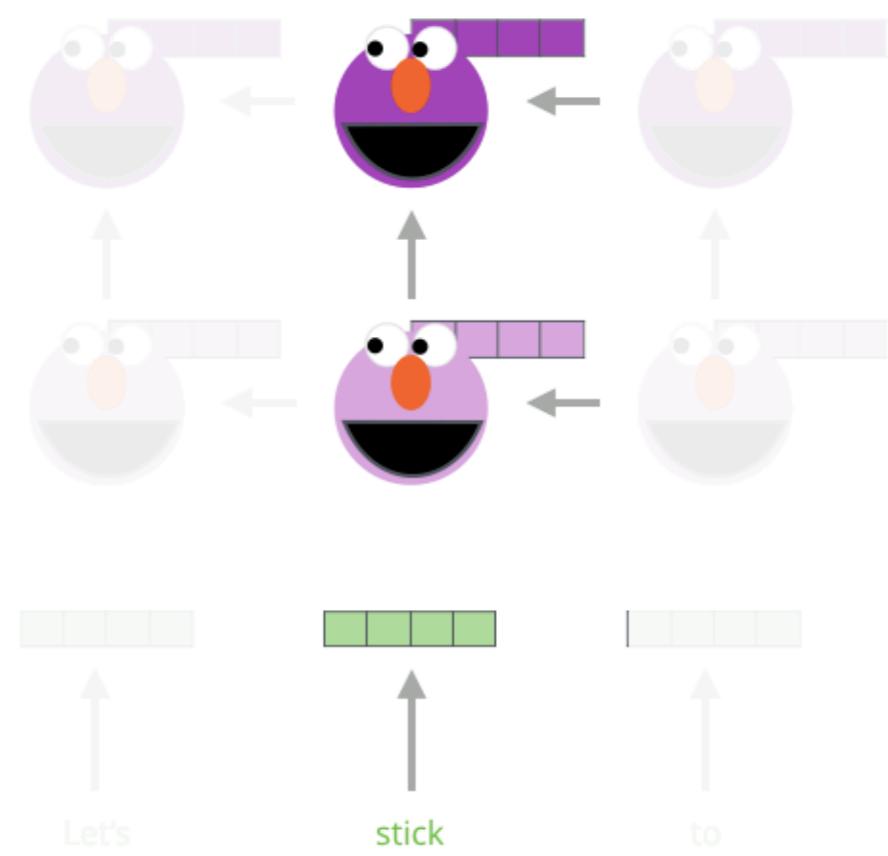
1- Concatenate hidden layers



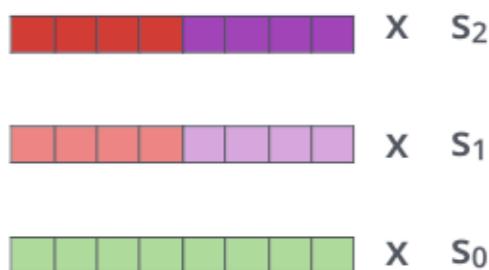
Forward Language Model



Backward Language Model



2- Multiply each vector by a weight based on the task



3- Sum the (now weighted) vectors



ELMo embedding of “stick” for this task in this context

# **Feature of NLP**

# Transformer XL



# Transformer XL

Чем плох Open AI Transformer?



# Transformer XL

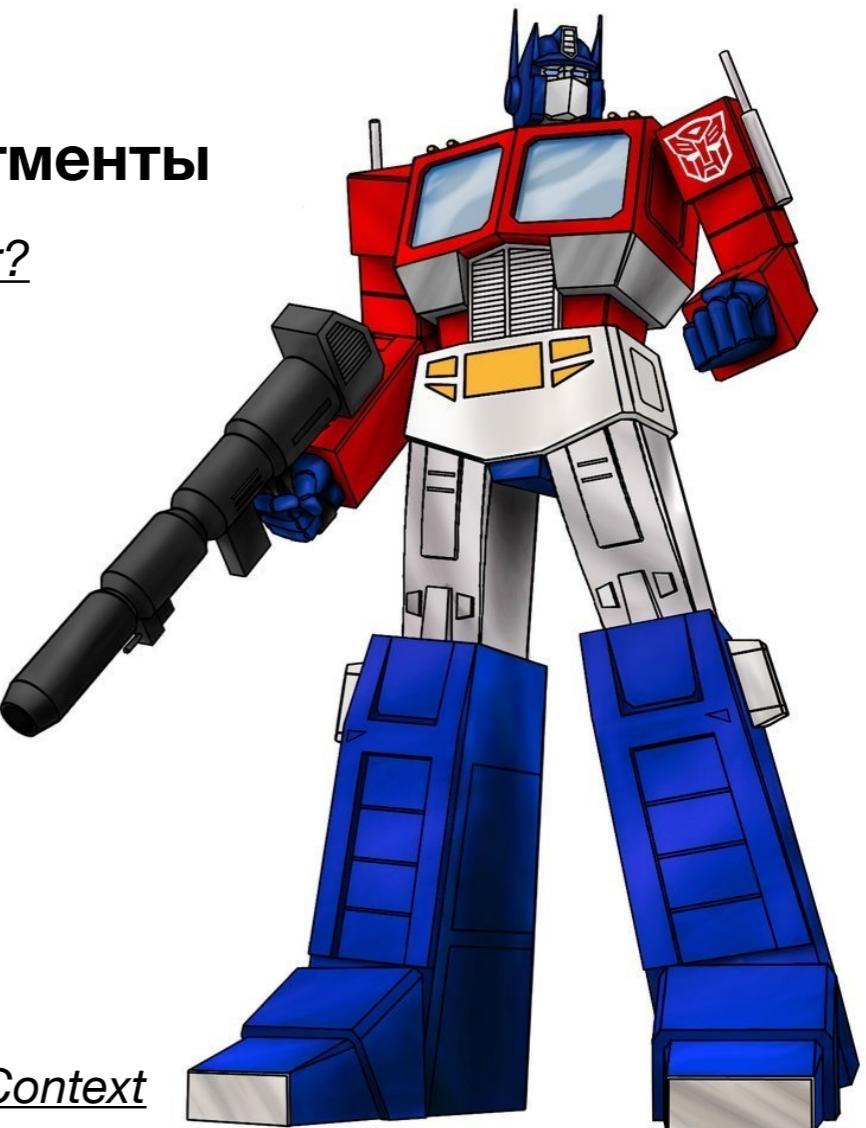
**Давайте разобьем наш текст на сегменты**



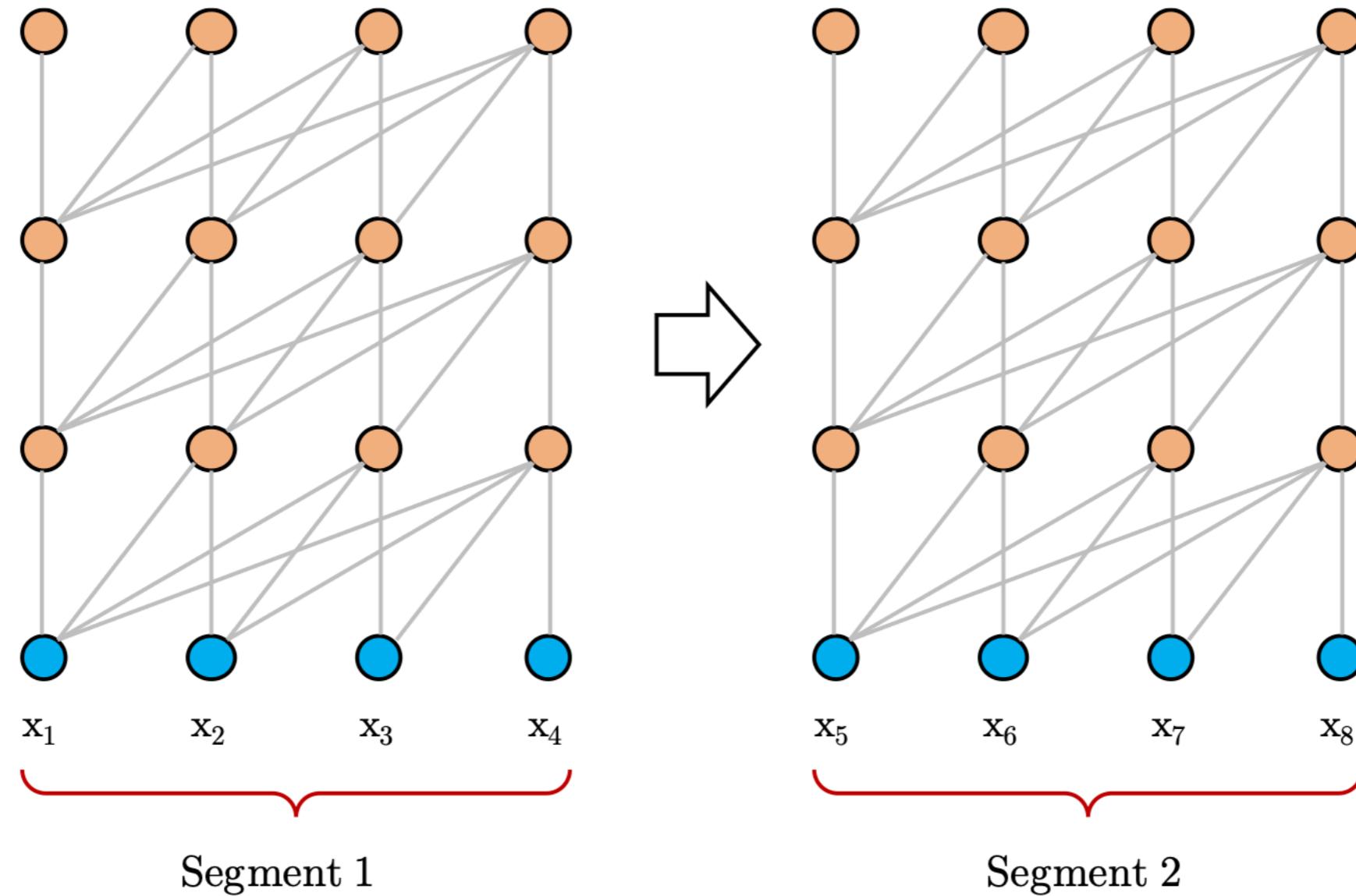
# Transformer XL

**Давайте разобьем наш текст на сегменты**

Как будет работать обычный transformer?

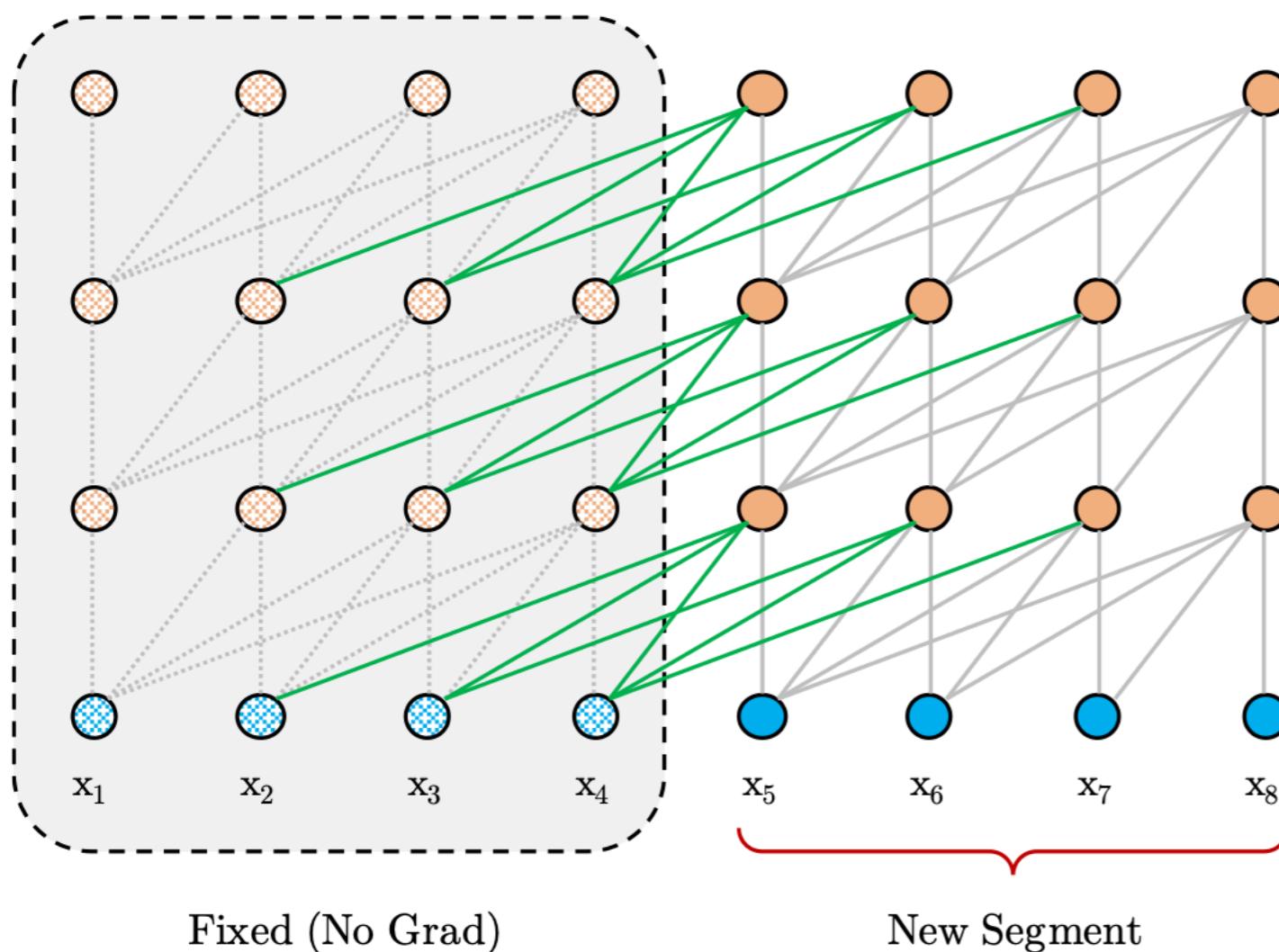


# Transformer XL



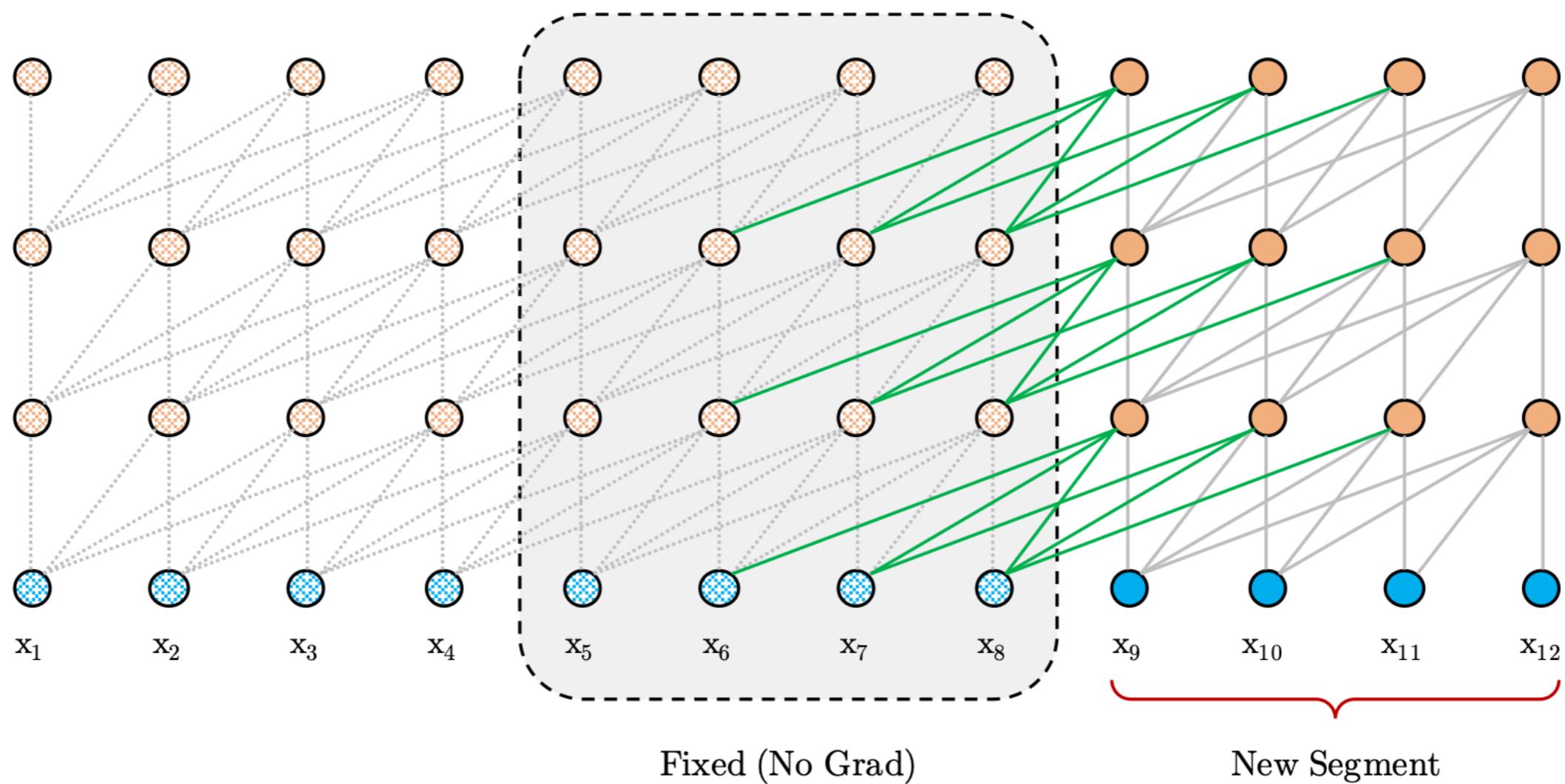
# Transformer XL

Segment-level recurrence:



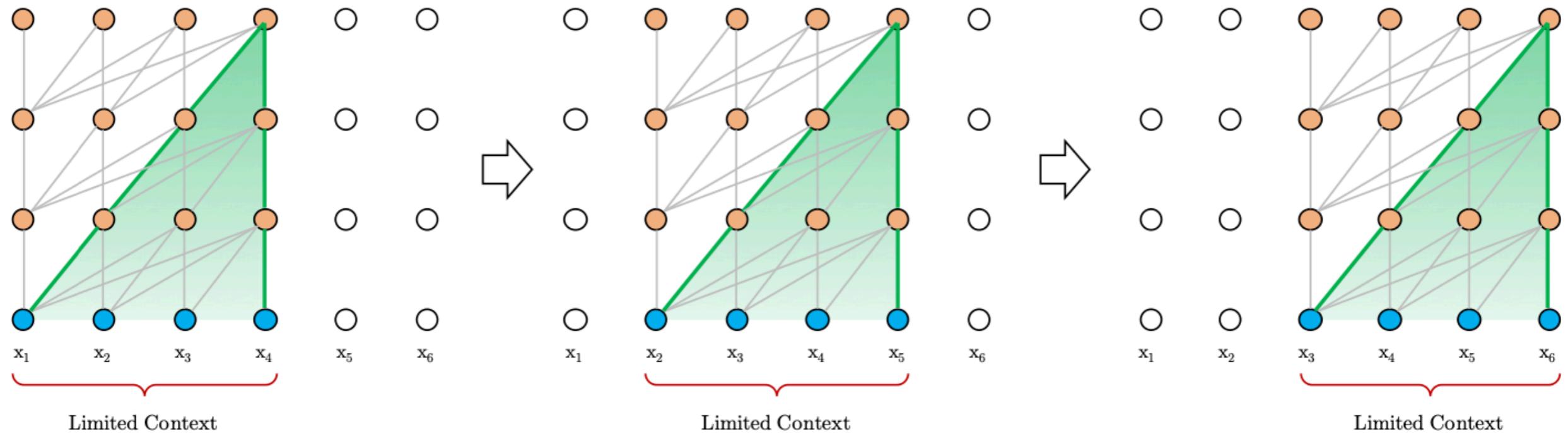
# Transformer XL

Segment-level recurrence:



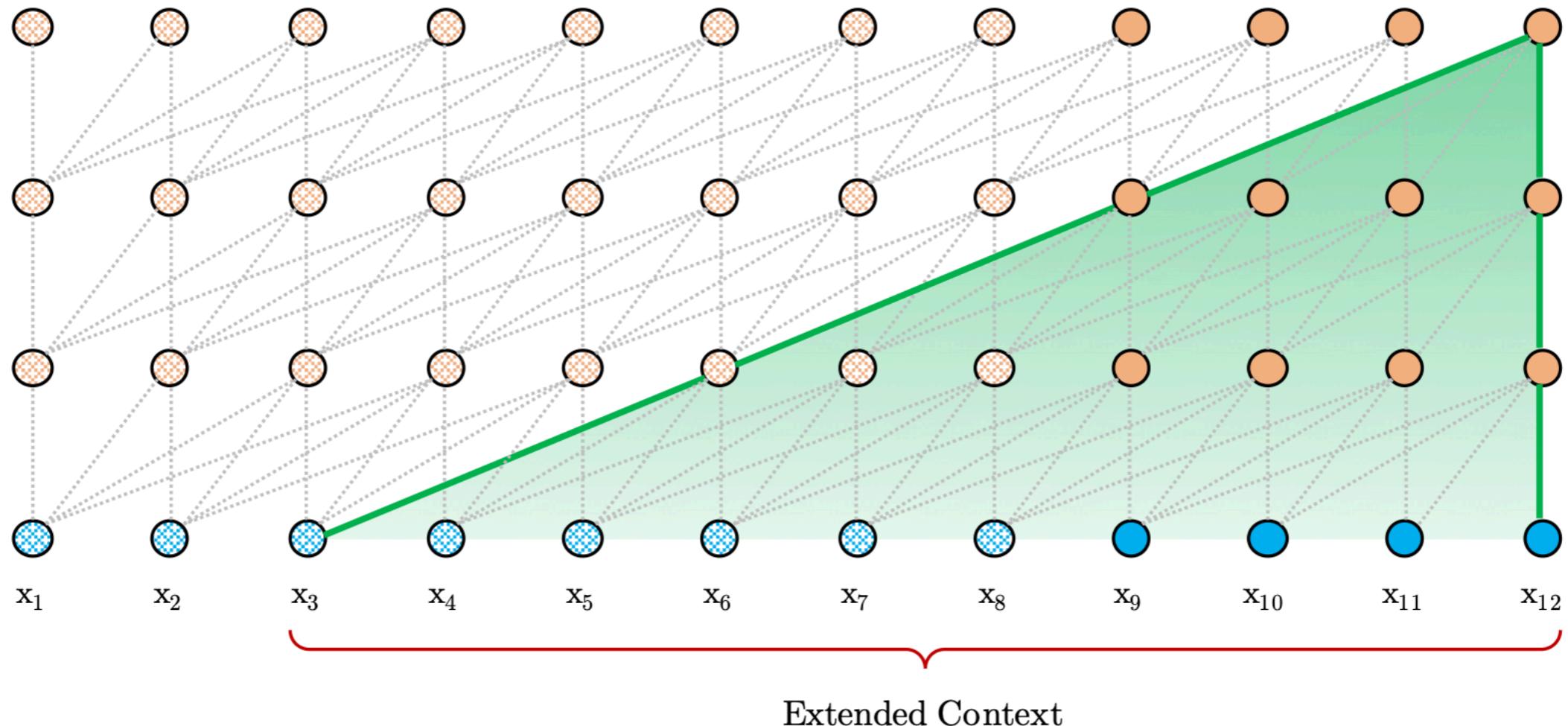
# Transformer XL

Evaluation phase of Vanilla transformer:



# Transformer XL

Evaluation phase of Transformer XL:



**SOTA GPT-2**

# SOTA GPT-2

- Transformer-based architecture
- trained to predict the **next** word
- 1.5 billion parameters
- Trained on 8 million web-pages



<https://openai.com/blog/gpt-2-6-month-follow-up/>

[https://github.com/ml-mipt/ml-mipt/blob/advanced/week05\\_BERT\\_and\\_LDA/Lecture\\_BERT\\_DIHT.pdf](https://github.com/ml-mipt/ml-mipt/blob/advanced/week05_BERT_and_LDA/Lecture_BERT_DIHT.pdf)

# GPT-2: fake news and hype

## Top stories



OpenAI built a text generator so good, it's considered too dangerous to release

TechCrunch

11 hours ago



Elon Musk's AI company created a fake news generator it's too scared to make public

BGR.com

9 hours ago



The AI That Can Write A Fake News Story From A Handful Of Words >

NDTV.com

2 hours ago