

Electronical Health Record  
Data Privacy Training Corpus **BERT** Ontology Construction Linear Activation  
Probability Distribution  
Meaningful Representation Character Codes  
Word Embedding king Drug Interaction  
**Natural Language Processing** queen  
Transformers Task spaCy Context  
word2vec Representation  
Named Entity Recognition  
gensim Hidden Layer Pre-trained Models Weight Matrix  
Kullback-Leibler Divergence  
Activation Function **Attention** Softmax Activation De-identification



<https://files.jansellner.net/NLPSeminar.pdf>

**dkfz.**




GERMAN  
CANCER RESEARCH CENTER  
IN THE HELMHOLTZ ASSOCIATION



Research for a Life without Cancer

- Named entity recognition
- Sentence similarity
- Family history extraction

The **transactivation T045** properties of the two **estrogen receptors** examined with different **ligands T103** in the context of an **estrogen** **ERalpha T116** and **ERbeta T116** were shown to **signal T067** i [1]

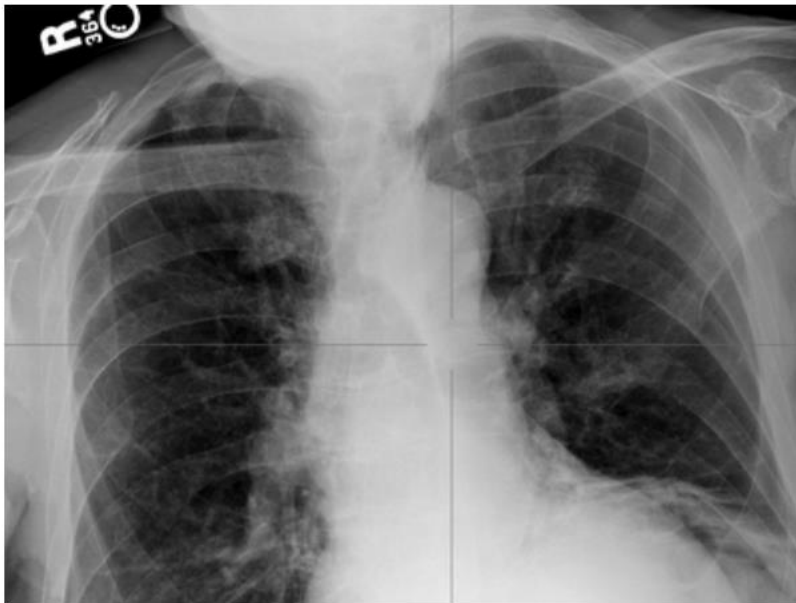
 Sister
<ul style="list-style-type: none"><li>• <b>305.0:</b> Alcohol abuse</li><li>• <b>433:</b> Occlusion and stenosis of precerebral arteries</li></ul>
 Aunt
<ul style="list-style-type: none"><li>• <b>346:</b> Migraine</li><li>• <b>366:</b> Cataract</li><li>• <b>433:</b> Occlusion and stenosis of precerebral arteries</li></ul>
 Mother
<ul style="list-style-type: none"><li>• <b>305.0:</b> Alcohol abuse</li><li>• <b>402.91:</b> With heart failure</li></ul>

[3]

S1: Mental: Alert and oriented to person place time and situation.  
S2: Feet:Neurological: He is alert and oriented to person, place, and time.

→ Similarity:  4.5/5 [2]

```
img = Image.open('example_image.png').convert('L')  
img.thumbnail((400, 500), Image.ANTIALIAS)  
img
```

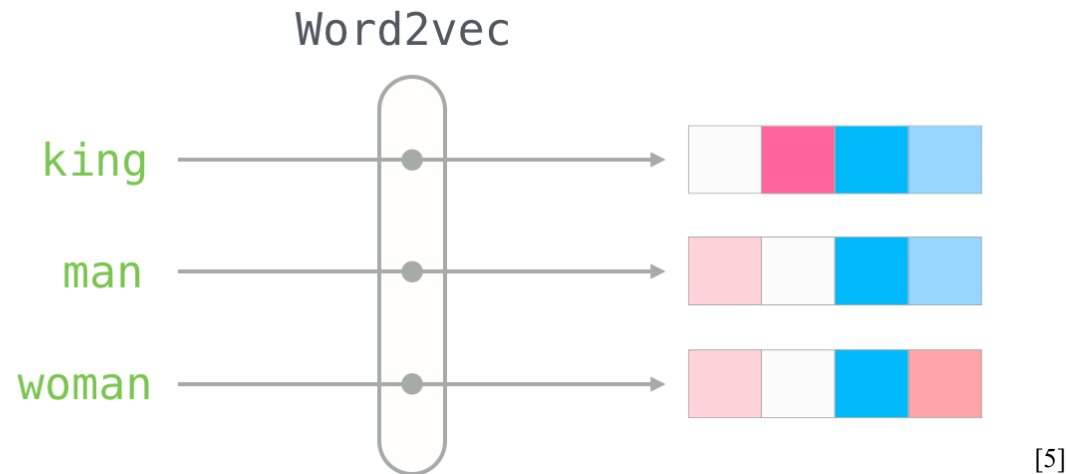


```
np.array(img)  
  
array([[ 0,  0,  0, ...,  0,  0,  0],  
       [ 0,  0,  0, ...,  0,  0,  0],  
       [ 0,  0,  0, ...,  0,  0,  0],  
       ...,  
       [ 58,  59,  62, ..., 124, 129, 133],  
       [ 58,  60,  65, ..., 124, 124, 128],  
       [ 60,  62,  68, ..., 127, 124, 125]], dtype=uint8)
```

There are new streaky left basal opacities which could represent only atelectasis; however, superimposed pneumonia or aspiration cannot be excluded in the appropriate clinical setting. There is only mild vascular congestion without pulmonary edema.

[4]

- Idea: map each word to a fixed-size vector from an embedding space



```
model['king']
```

```
array([ 0.50451,  0.68607, -0.59517, ..., -1.6106 , -0.64426, -0.51042],  
      dtype=float32)
```

```
model['queen']
```

```
array([ 0.37854 ,  1.8233  , -1.2648  , ..., -2.2839  ,  0.0092753,  
      -0.60284 ], dtype=float32)
```

- Understanding the meaning of a sentence is hard
  - Words have multiple meanings
  - Word compounds may alter the meaning
  - Coreference resolution
  - ...

The coin does not fit into the backpack because it is too small.

The coin does not fit into the backpack because it is too large.

The coin does not fit into the **backpack** because **it** is too small.



The **coin** does not fit into the backpack because **it** is too large.



The coin does not fit into the backpack because it is too small.

→ Die Münze passt nicht mehr in den Rucksack, weil er zu klein ist.

The coin does not fit into the backpack because it is too large.

→ Die Münze passt nicht mehr in den Rucksack, weil sie zu groß ist.



- BERT: language model developed by Google
  - Word embeddings aren't unique anymore; they depend on the context instead
  - Different architecture: the system has an explicit notion to model word dependencies → attention

## GLUE Benchmark (11 tasks)

Unsupervised pretraining followed by supervised finetuning



[6] J. Devlin et al. "BERT: Pre-training of deep bidirectional transformers for language understanding", arXiv:1810.04805, 2018 51

[7]

- Goal: model dependencies between words
- Idea: allow each word to pay attention to other words

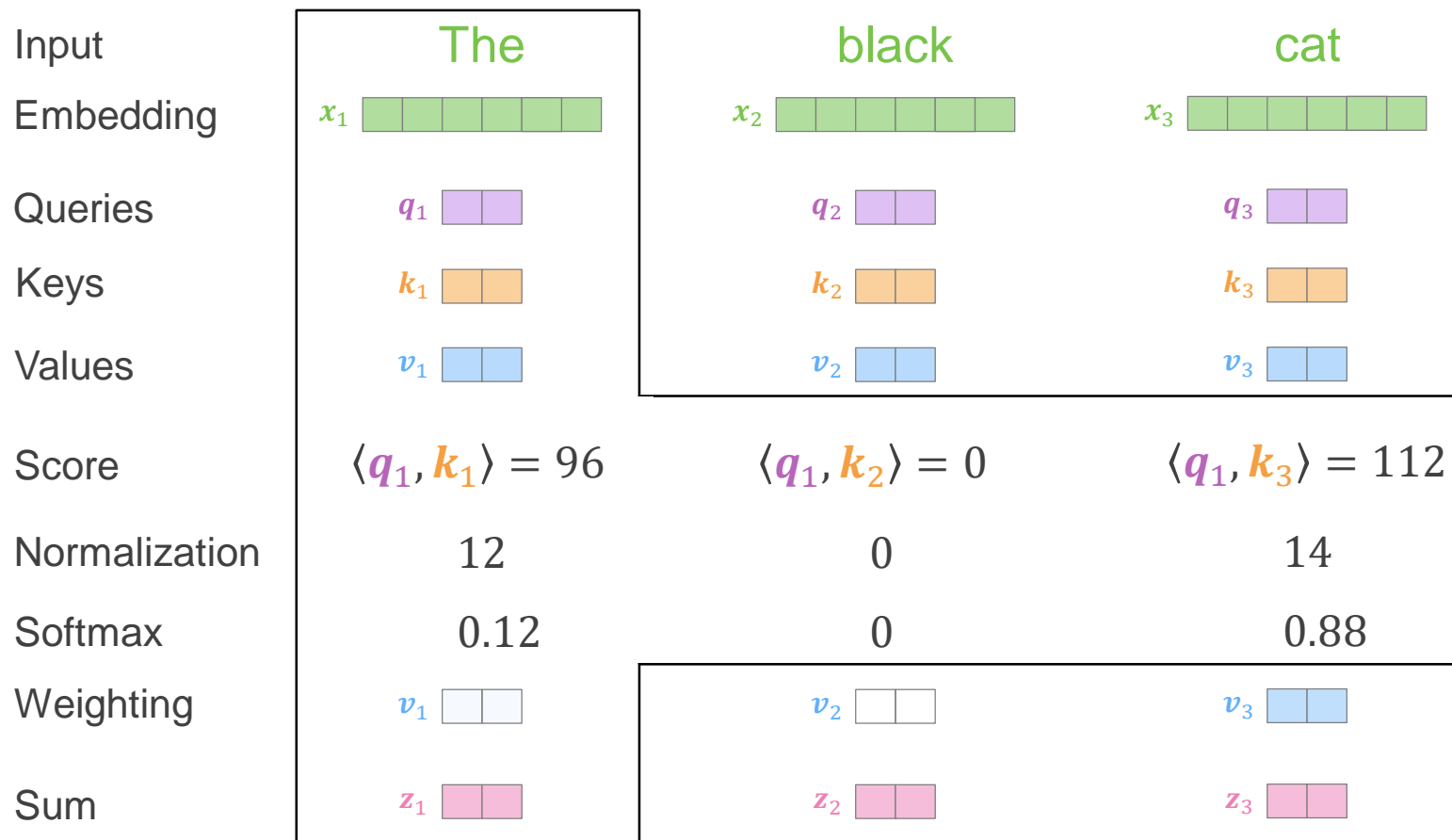


[12]

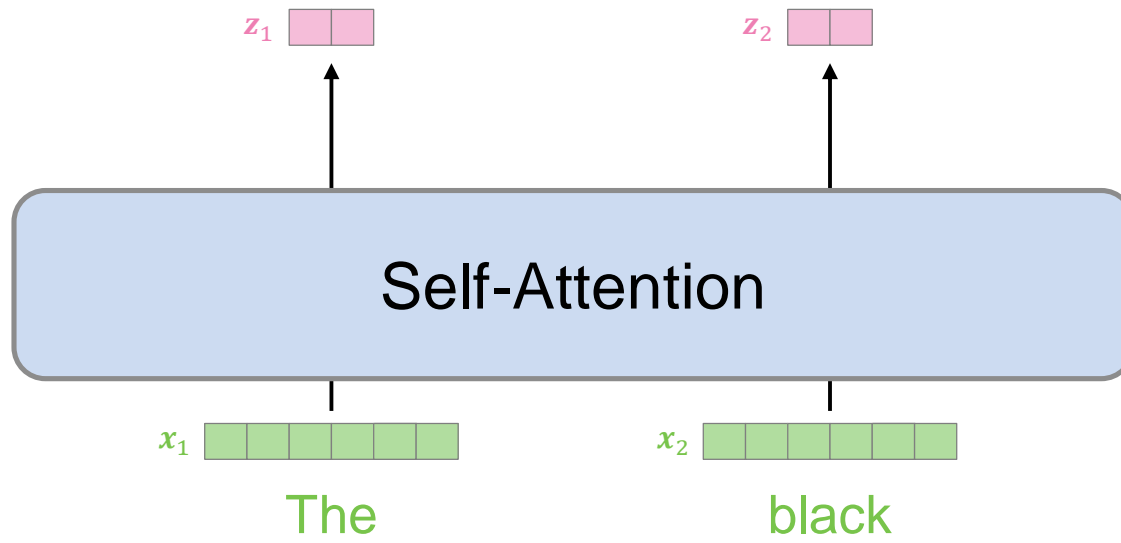
The black cat plays with the piano

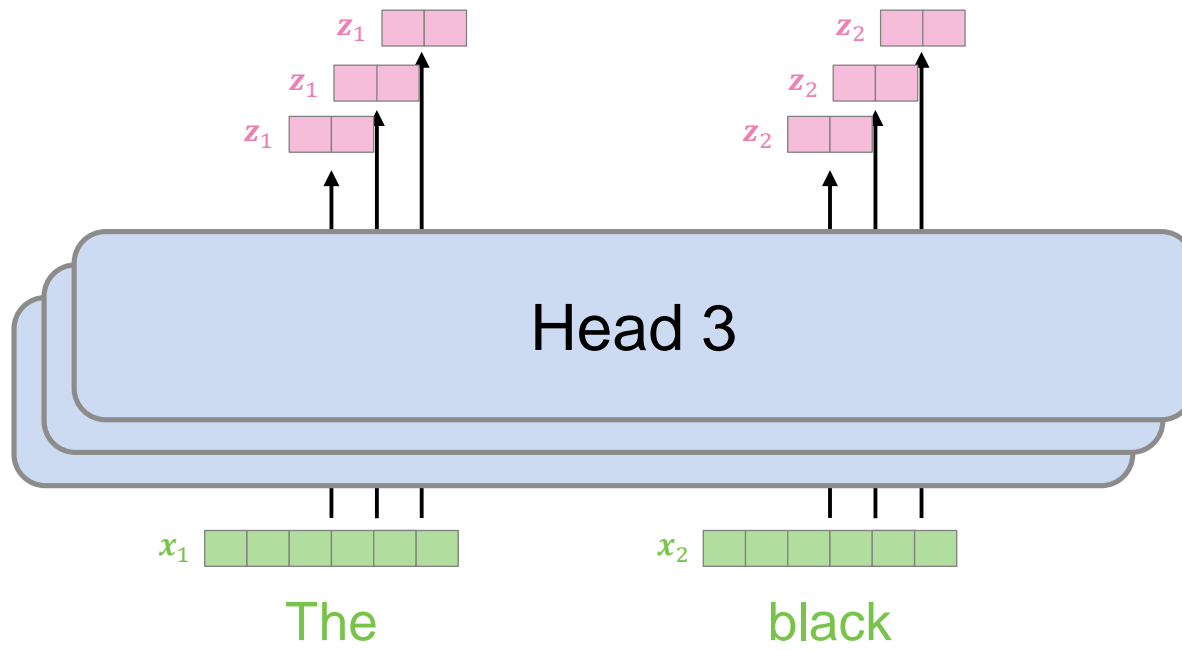
- “The” → “cat”: determiner-noun relationship
- “black” → “cat”: adjective-noun relationship
- “plays” → “with the piano”: verb-object relationship

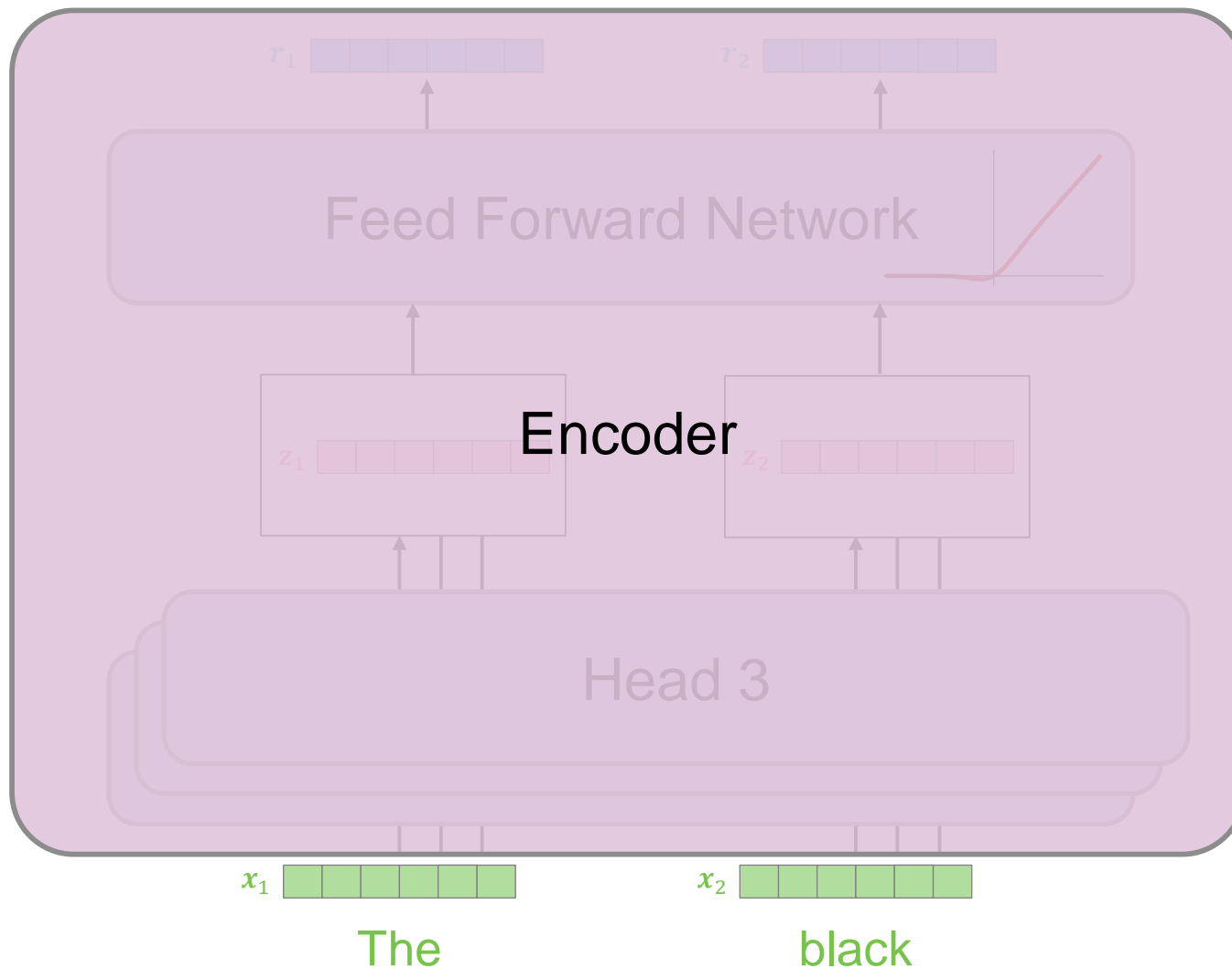
# How is Attention Calculated?

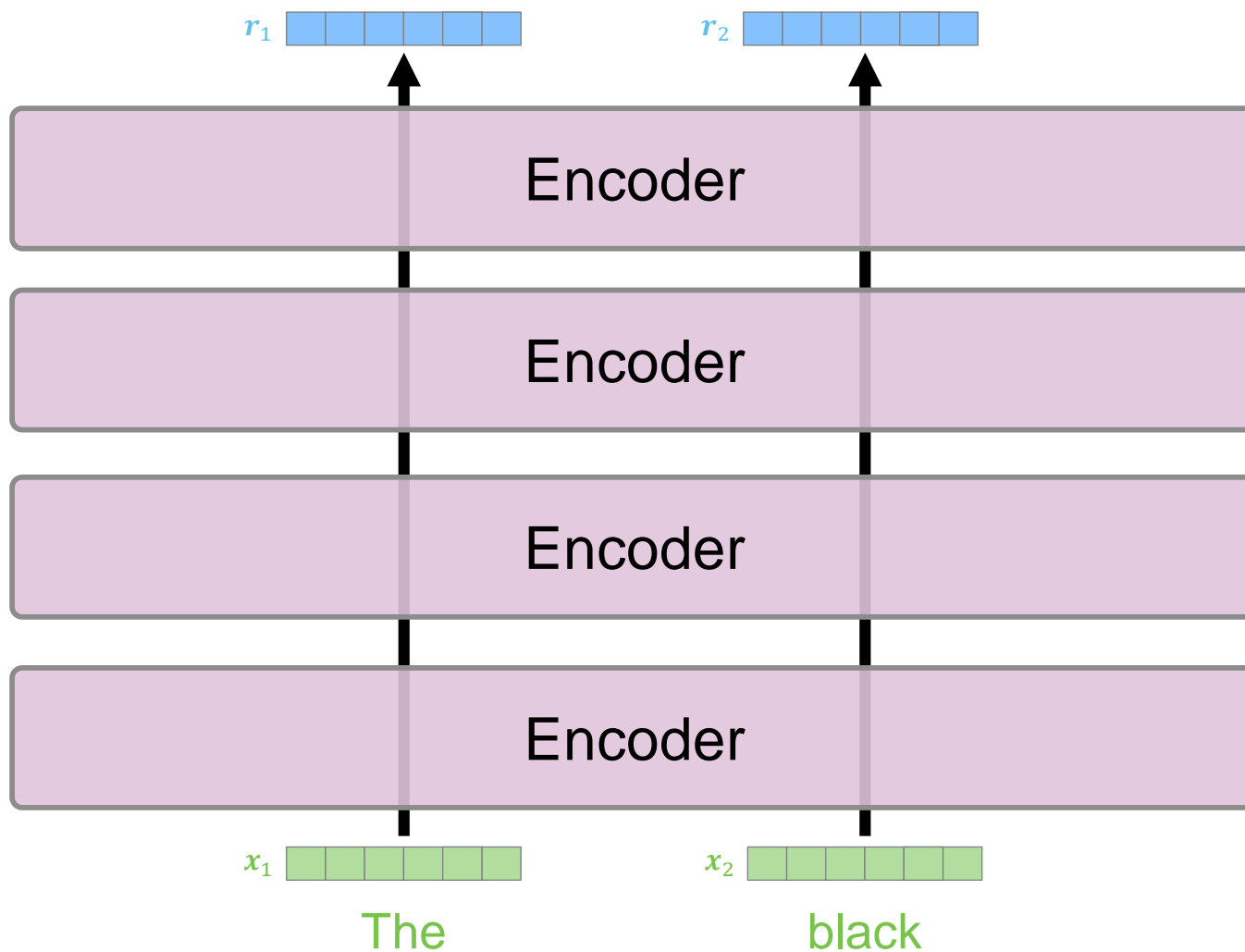


Based on [8]



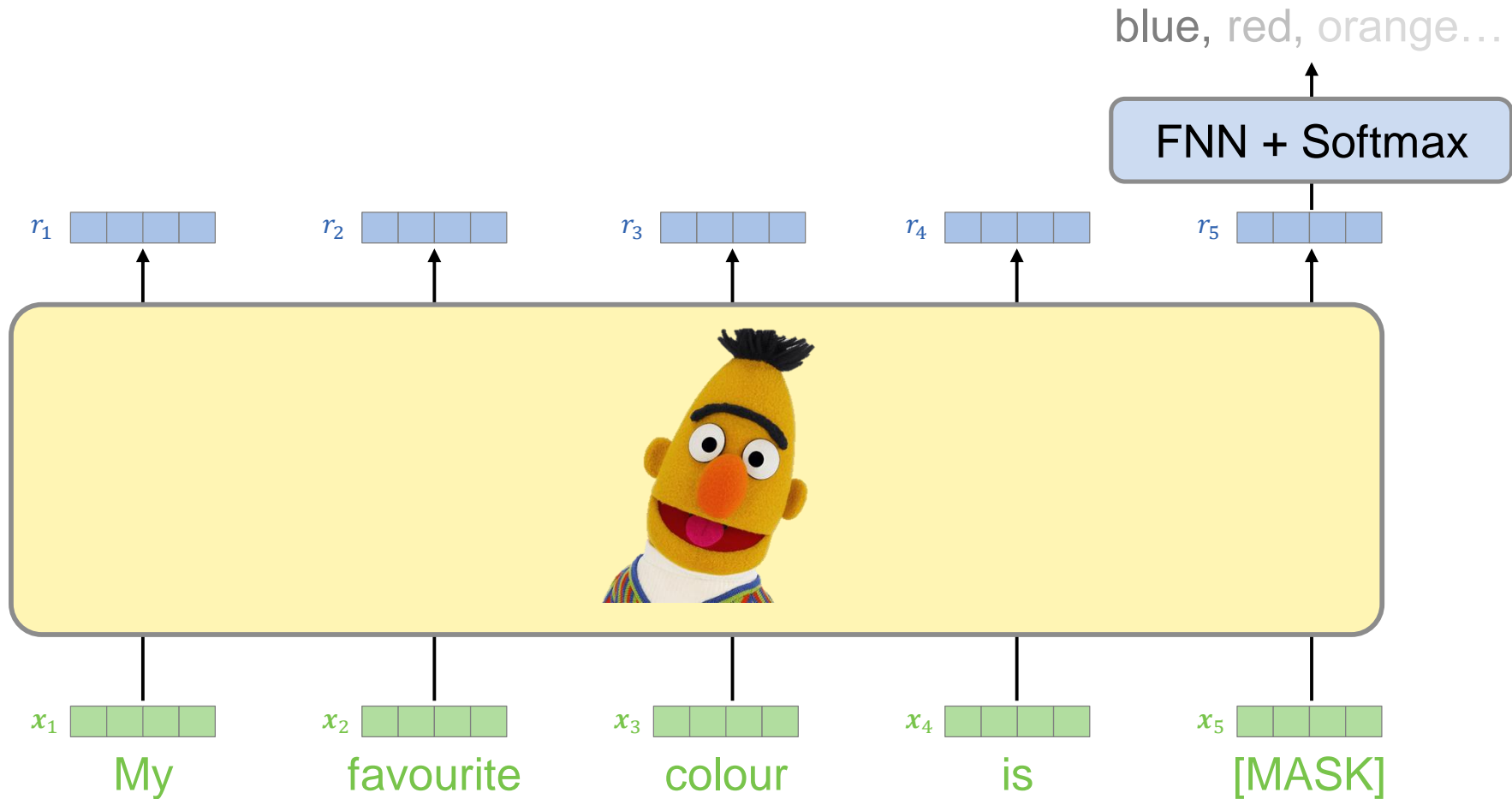


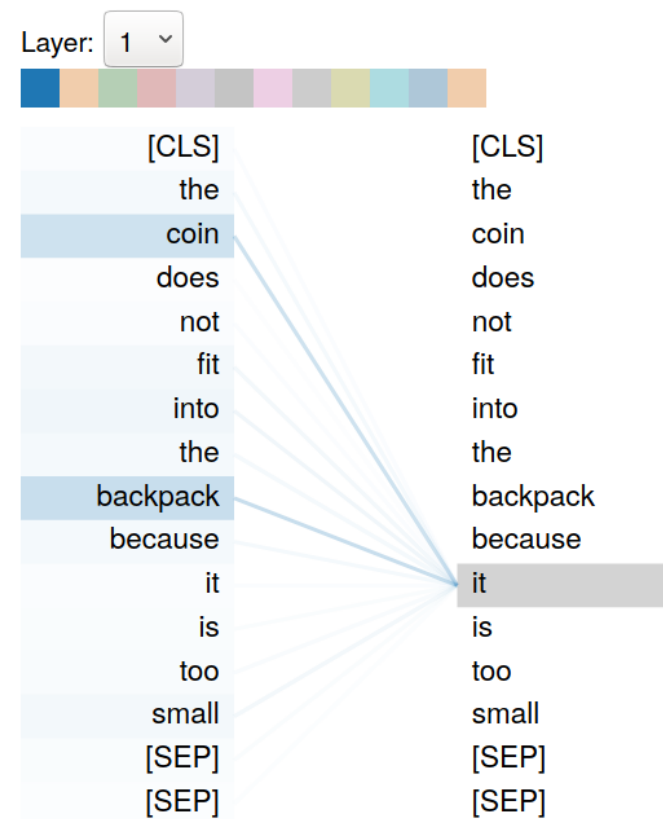
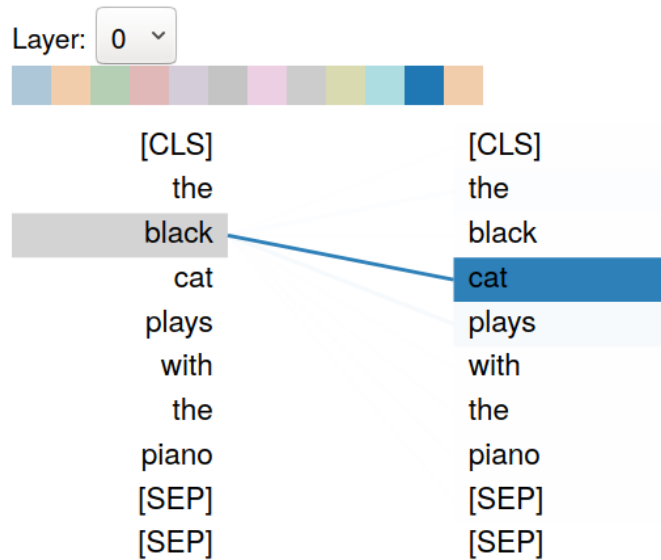




- Goal: BERT should get a basic understanding of the language
- Problem: not enough annotated training data available
- Idea: make use of the tons of unstructured data we have (Wikipedia, websites, Google Books) and define training tasks
  - Next sentence prediction
  - Masking

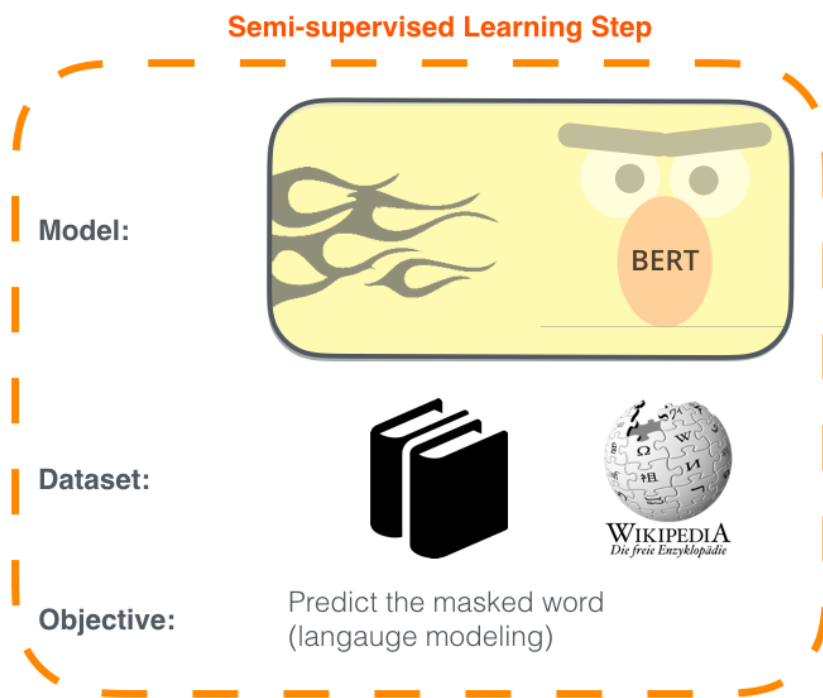




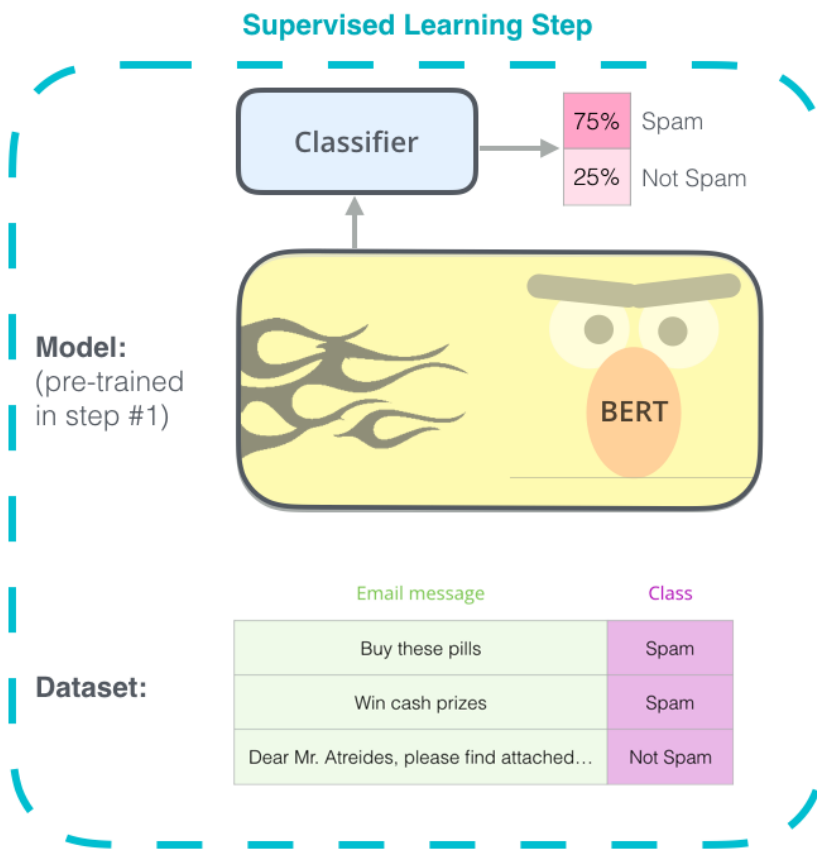


1 - **Semi-supervised** training on large amounts of text (books, wikipedia..etc).

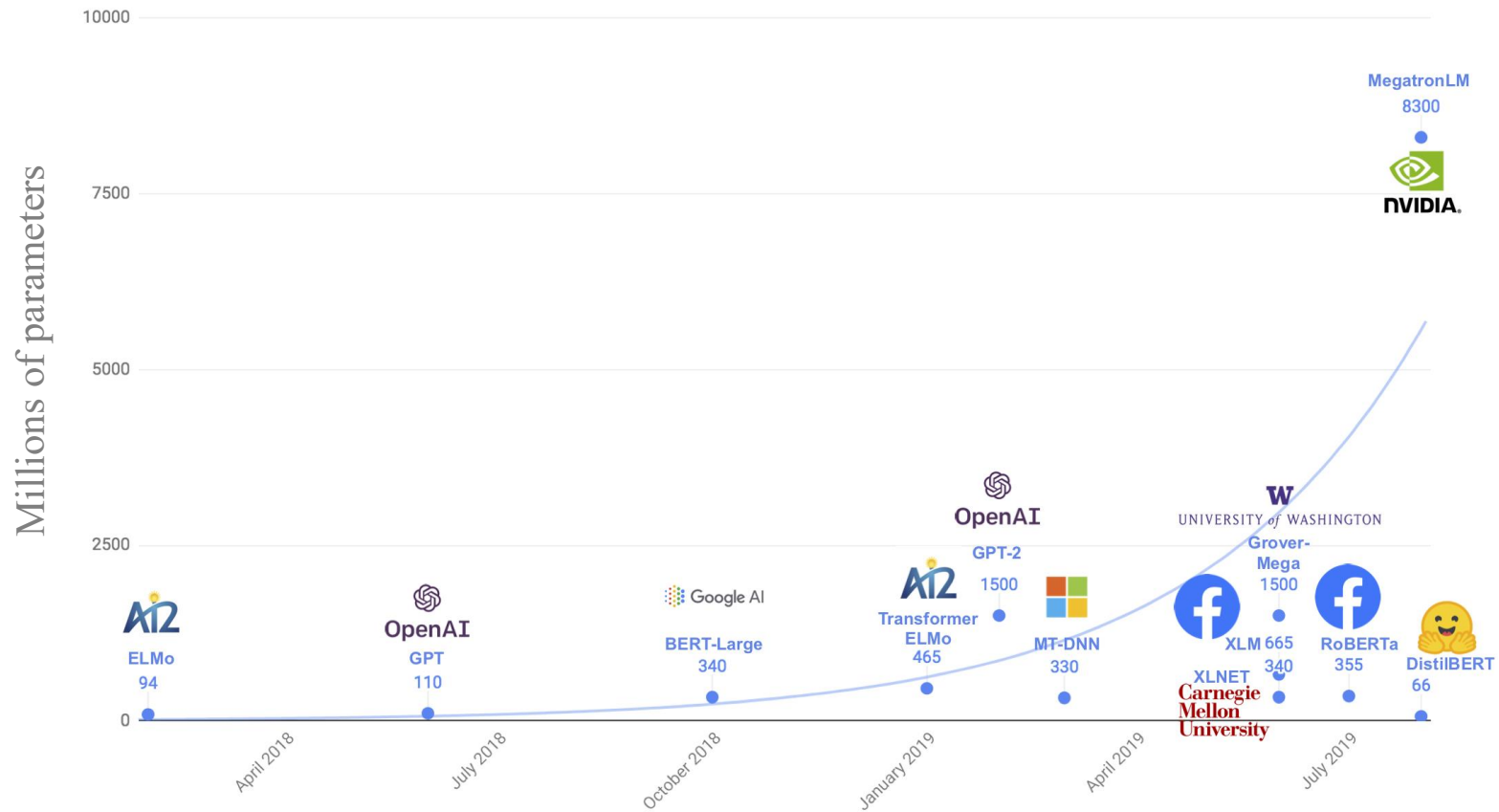
The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.



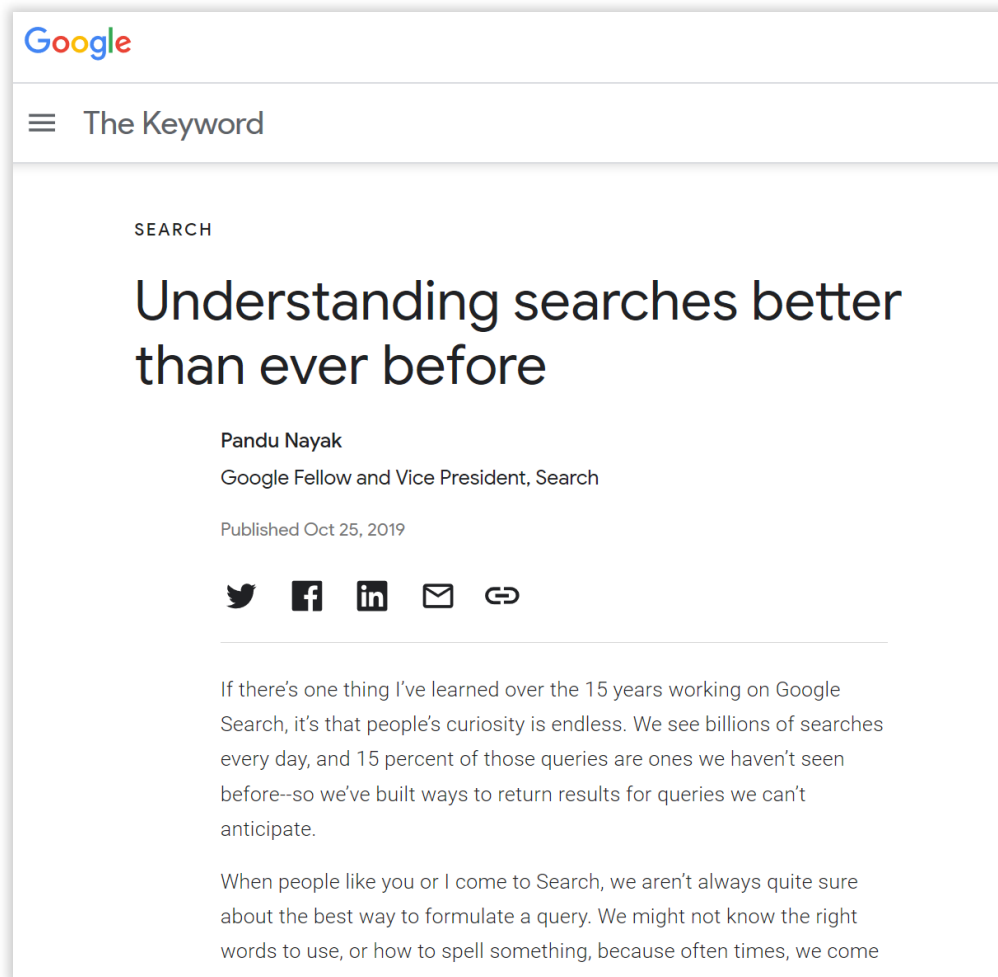
2 - **Supervised** training on a specific task with a labeled dataset.



[10]



[11]



<https://www.blog.google/products/search/search-language-understanding-bert/>

- Papers

- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. '[Attention Is All You Need](#)'. In NIPS, 2017.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. '[BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding](#)'. ArXiv:1810.04805 [Cs], 10 October 2018.

- Blogs

- <https://jalammar.github.io/>
- <https://medium.com/synapse-dev/understanding-bert-transformer-attention-isnt-all-you-need-5839ebd396db>
- <https://mccormickml.com/2019/05/14/BERT-word-embeddings-tutorial/>

- Implementation

- <https://github.com/huggingface/transformers>

- [1] [https://www.youtube.com/watch?v=2\\_HSKDALwuw&list=PLBmCuObd5An4UC6jvK\\_eSl6jCvP1gwXc](https://www.youtube.com/watch?v=2_HSKDALwuw&list=PLBmCuObd5An4UC6jvK_eSl6jCvP1gwXc)
- [2] [2019 n2c2 Shared-Task and Workshop - Track 1: n2c2/OHNLP Track on Clinical Semantic Textual Similarity](#)
- [3] Lewis, Neal, Gruhl, Daniel, Yang, Hu. '[Extracting Family History Diagnosis from Clinical Texts](#)'. In BICoB, 2011.
- [4] Johnson, Alistair E W, Pollard, Tom J, Berkowitz, Seth, Greenbaum, Nathaniel R, Lungren, Matthew P, Deng, Chih-ying, Mark, Roger G, Horng, Steven. '[MIMIC-CXR: A large publicly available database of labeled chest radiographs](#)'. [arXiv preprint arXiv:1901.07042](#), 2019
- [5] <https://jalammar.github.io/illustrated-word2vec/>
- [6] [https://twitter.com/seb\\_ruder/status/1070470060987310081/photo/3](https://twitter.com/seb_ruder/status/1070470060987310081/photo/3)
- [7] <https://mc.ai/how-to-fine-tune-and-deploy-bert-in-a-few-and-simple-steps-to-production/>
- [8] <https://jalammar.github.io/illustrated-transformer/>
- [9] <https://github.com/jessevig/bertviz>
- [10] <https://jalammar.github.io/illustrated-bert/>
- [11] <https://medium.com/huggingface/distilbert-8cf3380435b5>
- [12] <https://medium.com/synapse-dev/understanding-bert-transformer-attention-isnt-all-you-need-5839ebd396db>