Modern NLP in claims management

Leandro von Werra, Data Scientist @ Swiss Mobiliar

Acturial Data Science Après-Midi

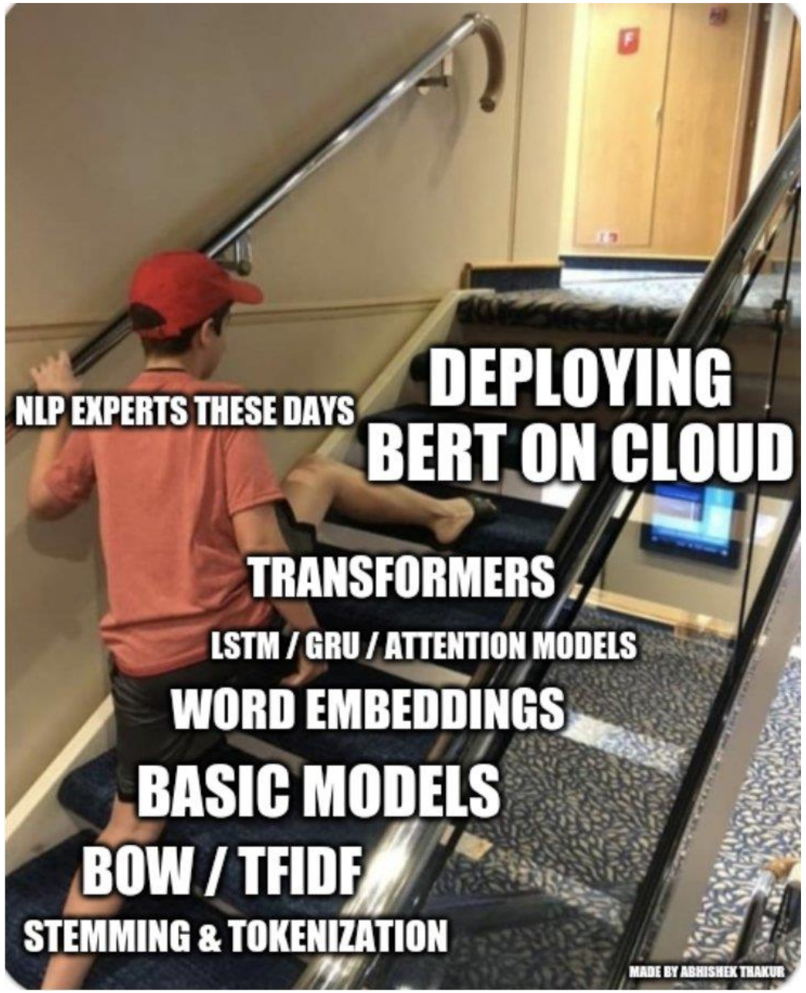
November 4, 2020

Contents of this talk: three eras

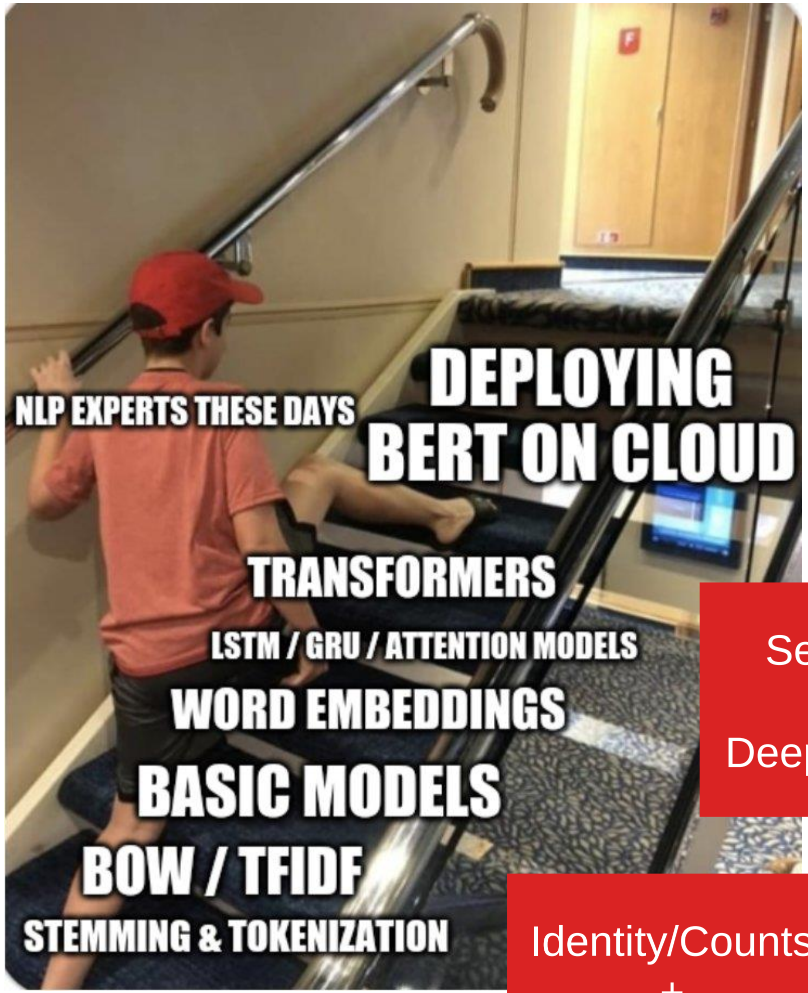
2013 | **2017**

Identity/Counts + Classical ML	Semantics + Deep Learning	Contexts + Transfer Learning

Contents of this talk: three eras



Contents of this talk: three eras

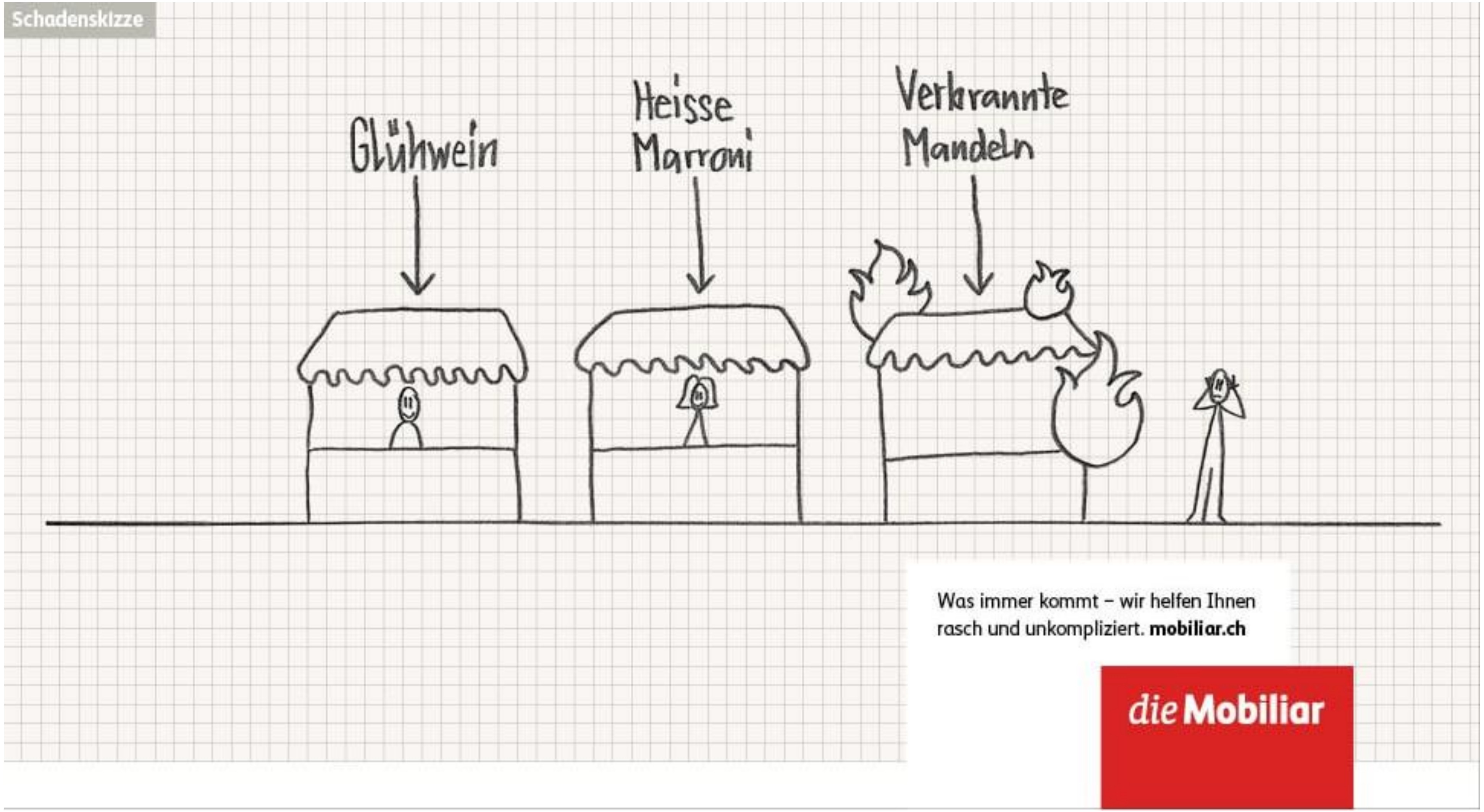


Contexts
+
Transfer Learning

Semantics
+
Deep Learning

Identity/Counts
+
Classical ML

Claims at Swiss Mobiliar: Expectation



Claims at Swiss Mobiliar: Reality

«Liebe Mobiliar,
ein Marder hat bei meinem VW Golf einige Stromkabel durchgebissen und ich musste meinen Wagen in die Werkstatt bringen ☹»

Part 1:

Identity/Counts

+

Classical ML

Claims at Swiss Mobiliar: Reality

«Liebe Mobiliar,
ein Marder hat bei meinem VW Golf einige Stromkabel durchgebissen und ich musste meinen Wagen in die Werkstatt bringen ☹»

Schadenfall?

Teilfall?

Auslöser?

→ Classification task!

Text representation: Identity & count encodings

How can we feed textual data to a numerical model?

One-hot encodings:

[0 0 1 0 0 0 0 0] [1 0 0 0 0 0 0 0]

[1 0 0 0 0 0 0 0] [0 0 0 0 1 0 0 0] [0 0 0 0 0 0 0 1]

The quick brown fox jumps over the lazy dog.

[0 1 0 0 0 0 0 0] [0 0 0 0 0 1 0 0]

[0 0 0 1 0 0 0 0] [0 0 0 0 0 0 1 0]

(size of the vectors = number of words in the vocabulary of the corpus)

How can we encode a whole sentence/document?

Count encodings:

[1 0 0 0 0 0 0 0] + [0 1 0 0 0 0 0 0] + [0 0 1 0 0 0 0 0] + … = [2 1 1 1 1 1 1 1]

Encodings: Pros and Cons

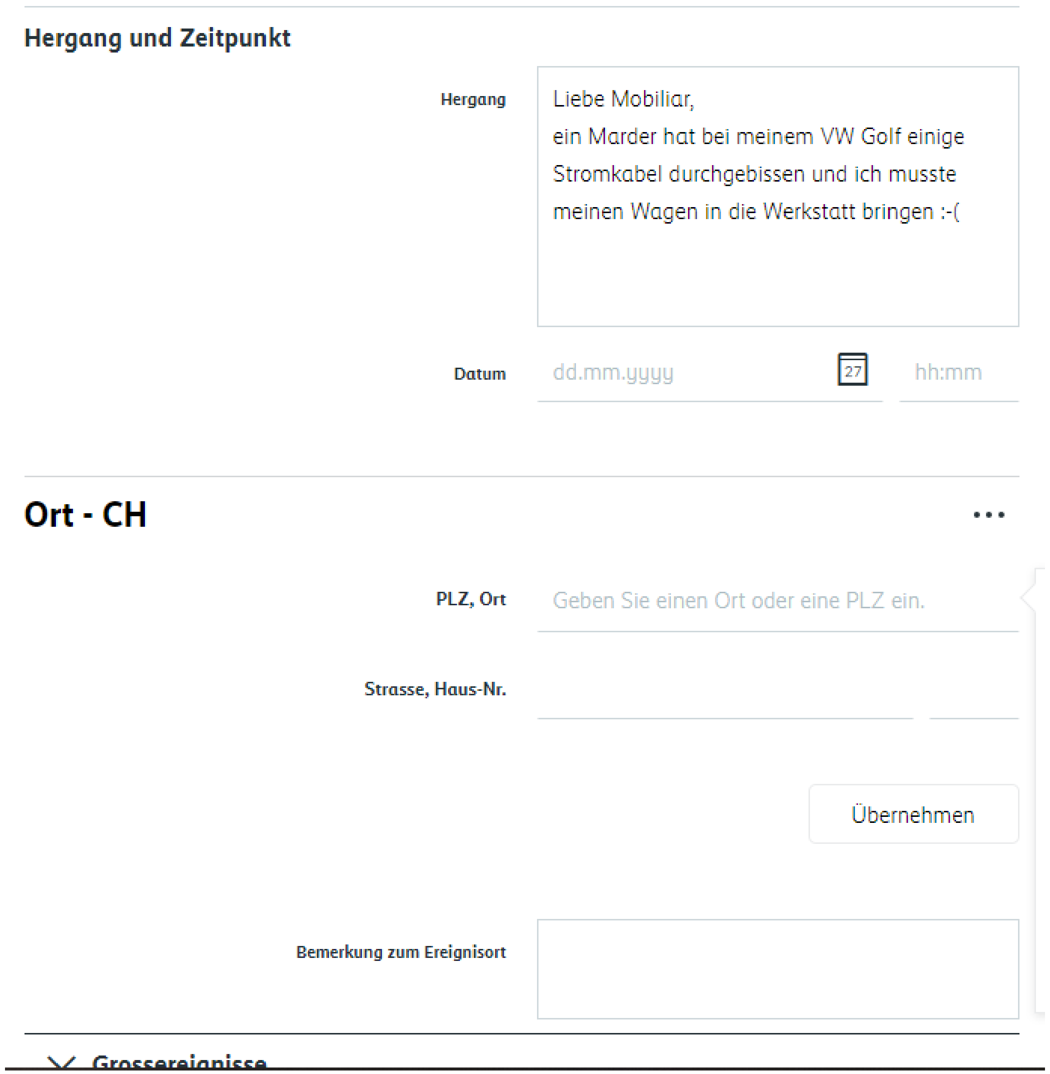
Pros

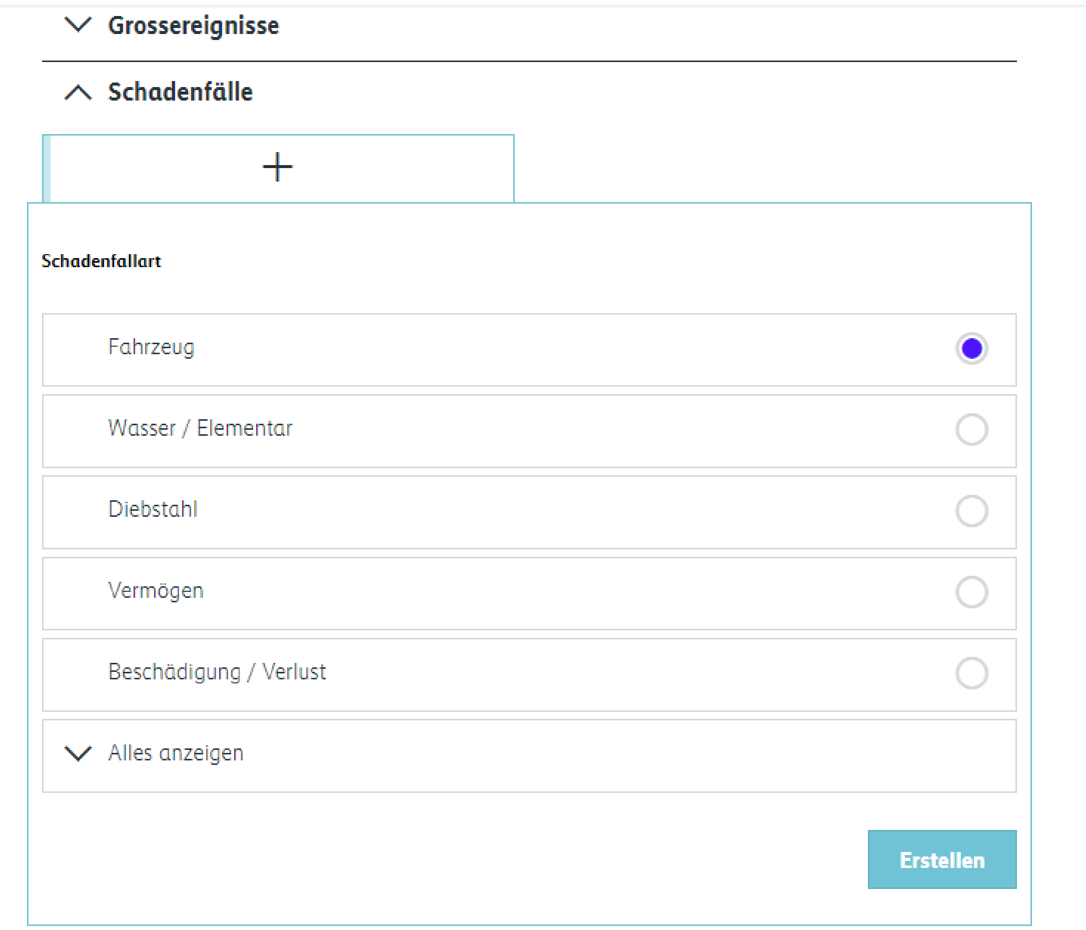
- Simple to construct (vocab + count)
- Feature for ML models (SVM, Naïve Bayes, Neural Networks etc.)
- Easy decoding

Cons

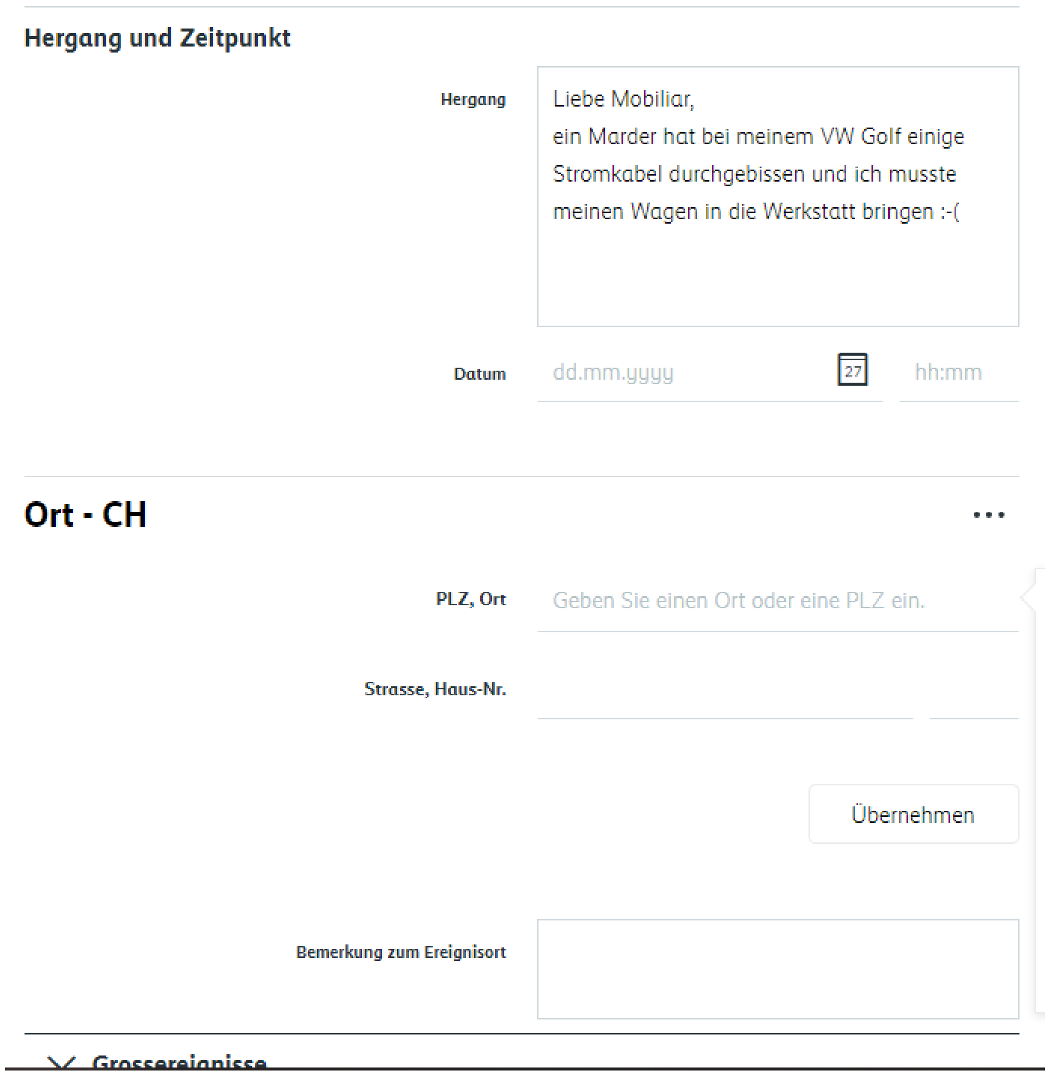
- Large feature space (O(10^6))
- Loss of sequential information (→n-grams, RNNs)

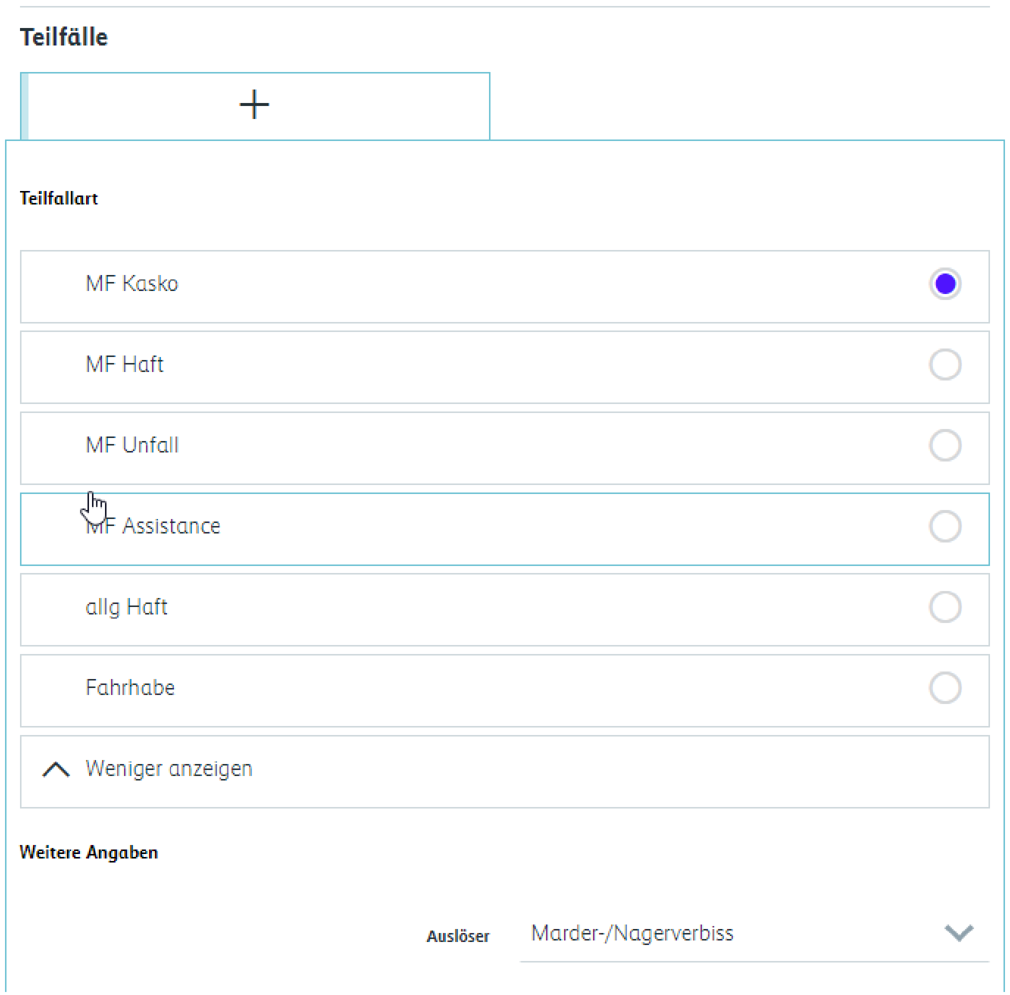
In production: n-grams + Naïve Bayes





In production: n-grams + Naïve Bayes





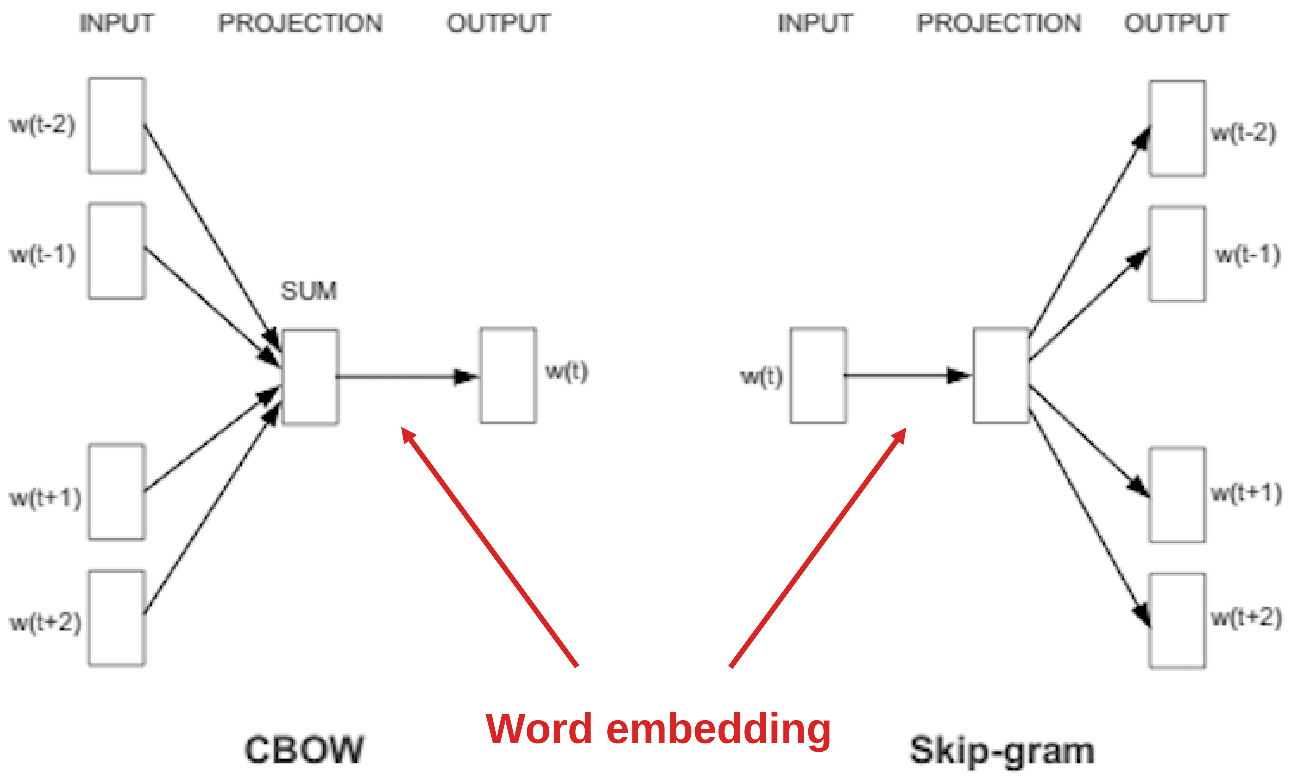
Part 2:

Semantics

+

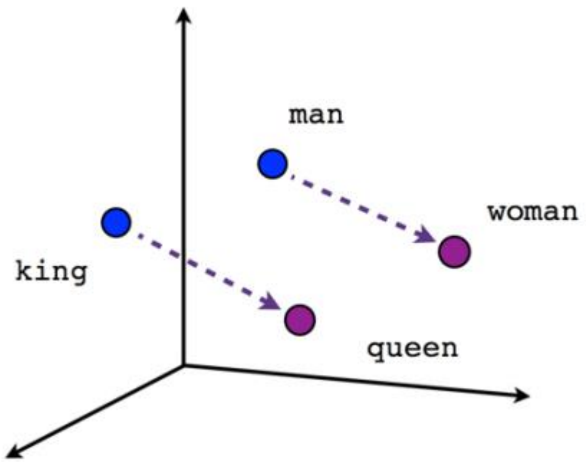
Deep Learning

Semantic word embeddings: word2vec

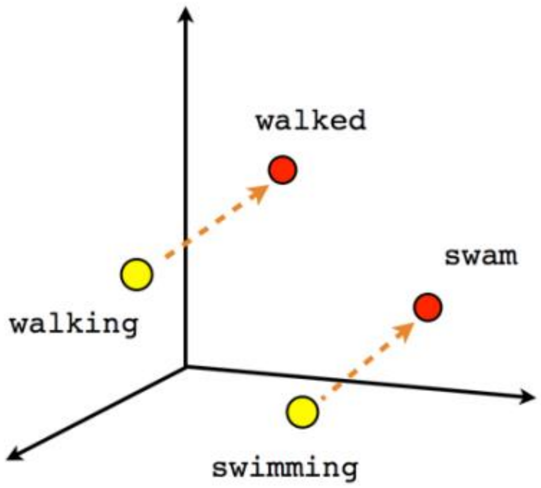


https://towardsdatascience.com/understanding-word2vec-embedding-in-practice-3e9b8985953

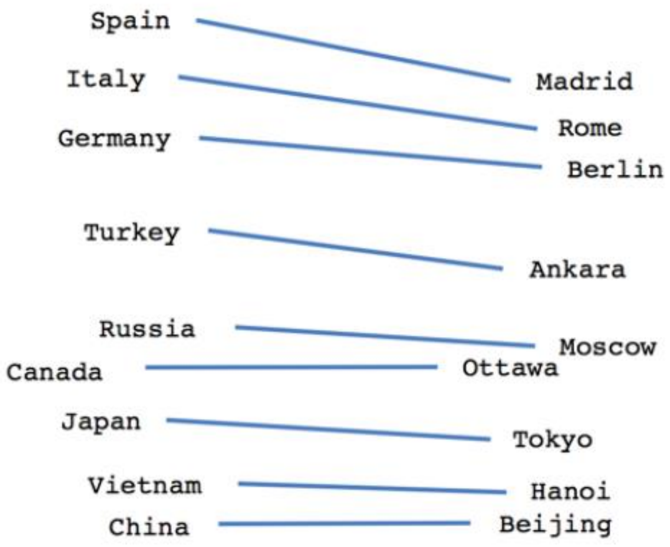
Vector geometry with word embeddings



Male-Female



Verb tense

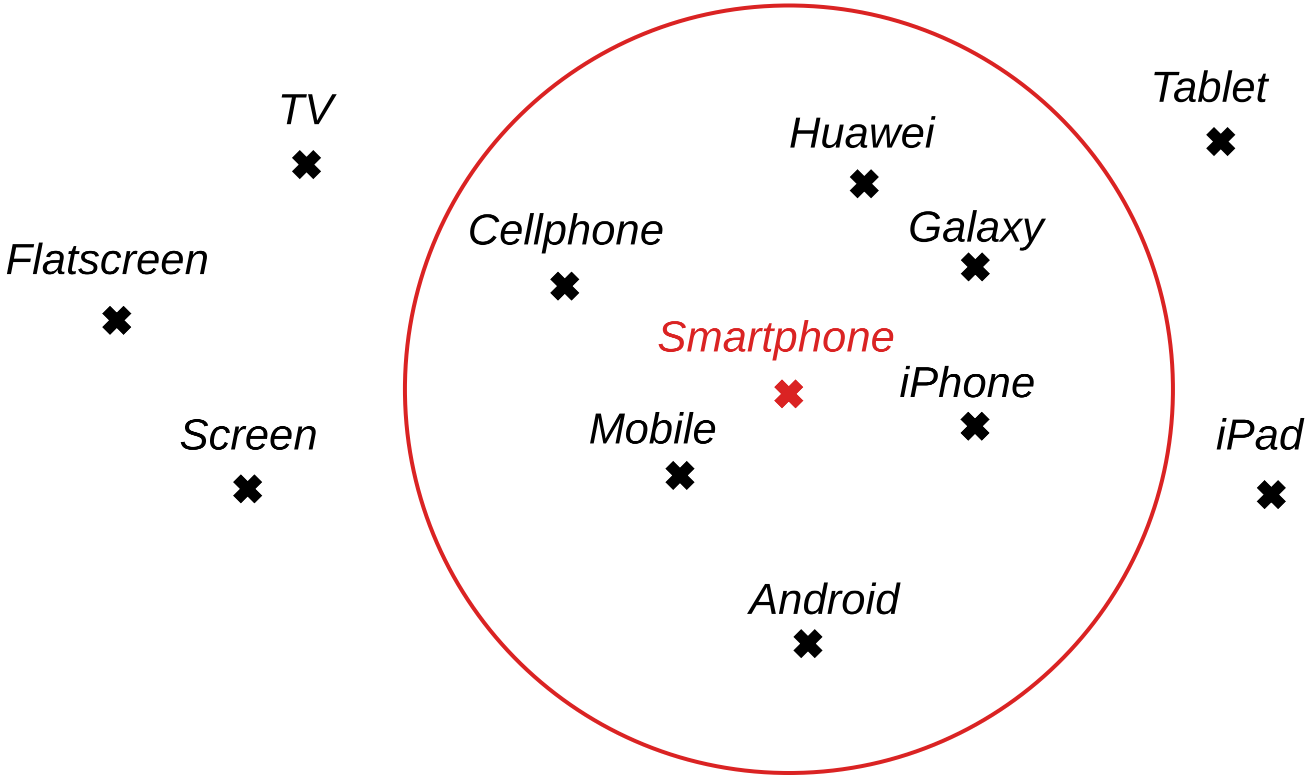


Country-Capital

Similarity search for trends

How many claims for Smartphones do we receive?

Keyword: *Smartphone*
Synonyms: *Cellphone, iPhone, Huawei, Mobile, Samsung Galaxy etc.*



Similarity search for trends

ANALYSE VON SCHADENHERGÄNGEN

Sucht nach ähnlichen Schadensmeldungen innerhalb einer ausgewählten Schadenbranche und Schadenursache und zeigt die entsprechende Verteilung über die vergangenen Jahre an.

Suchbegriffe:

boden wasser

Schadenbranche:

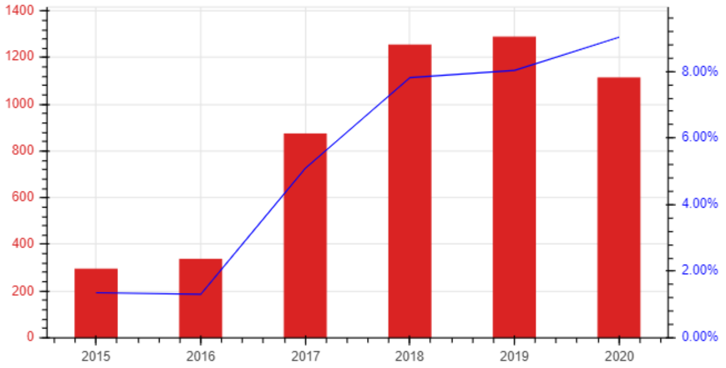
Bauherren-Haftpflicht

Schadenursache:

AGM Fremde

Similarity: **0.90 .. 1**

Ähnliche Schadenfälle suchen



Word embeddings: Pros and Cons

Pros

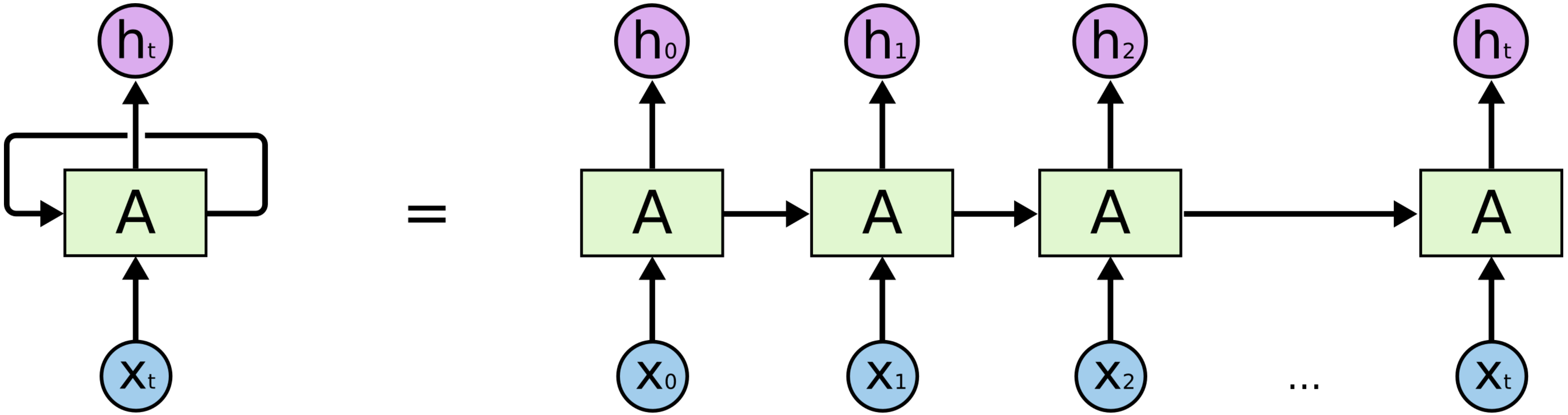
- Partly encode meaning
- Dense input features
- Jumpstart model training
- Similarity measure between words/sentences

Cons

- Bias issues
- Don't solve sequential problem
- Static representation (Apple (company) and apple (fruit))

Recurrent Neural Networks

Retain sequential information with RNNs:



Require a lot of training data and are hard to scale!

https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Part 3:

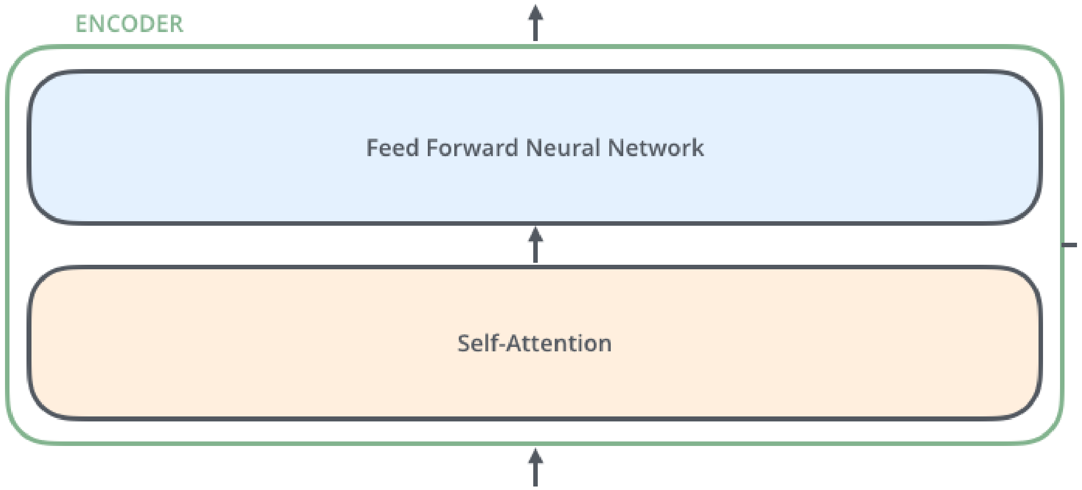
Context

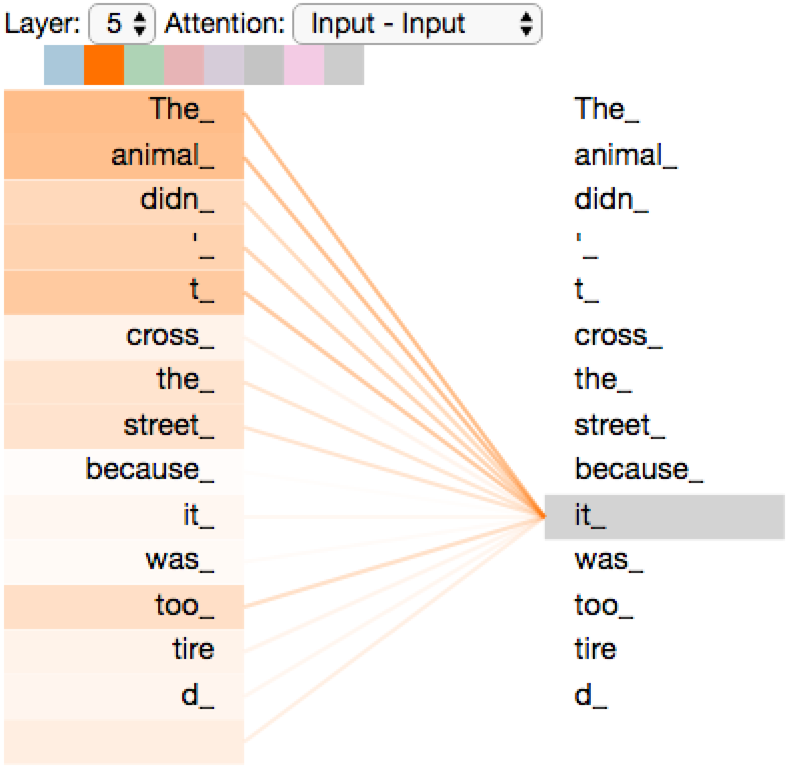
+

Transfer Learning

Transformers: Attention is all you need

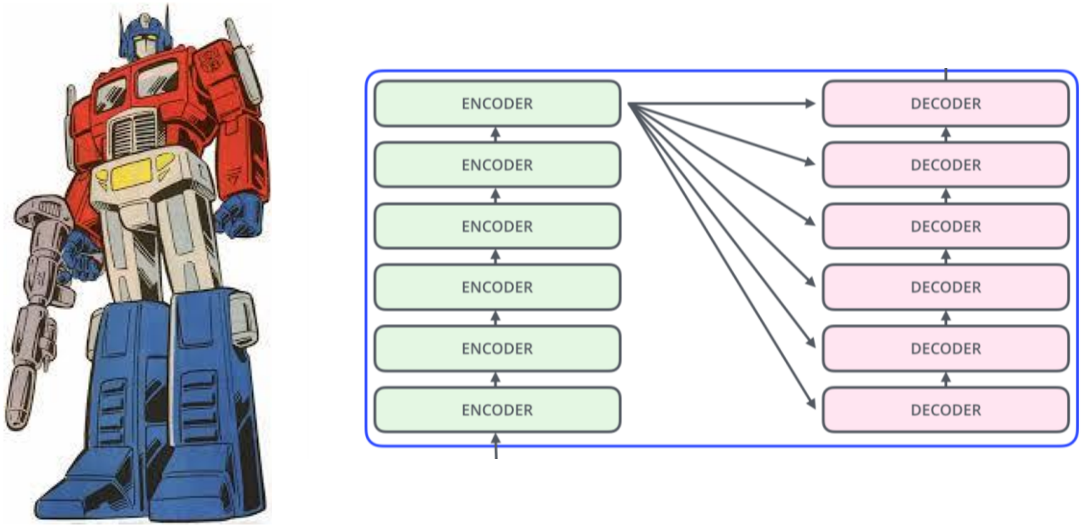
Feedforward Networks + Attention
→ No recurrence neccessary

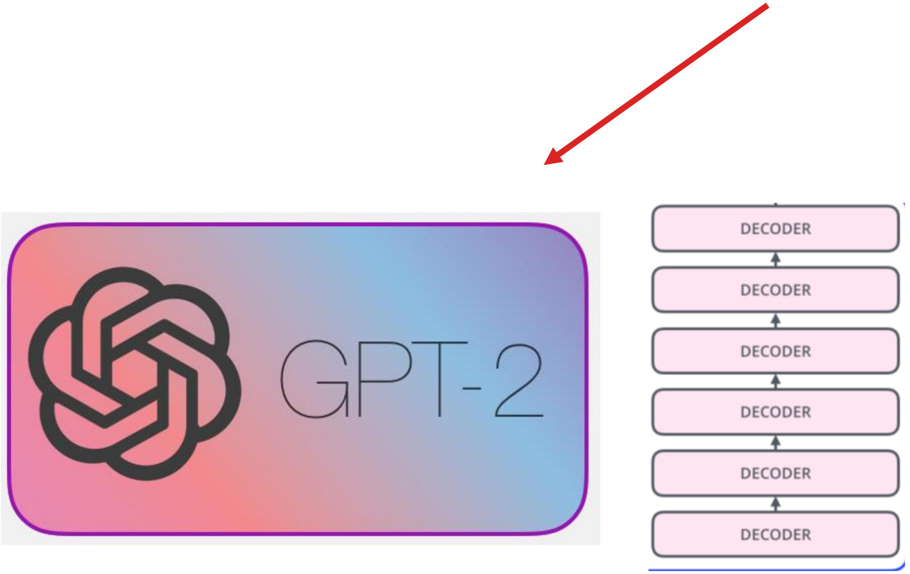


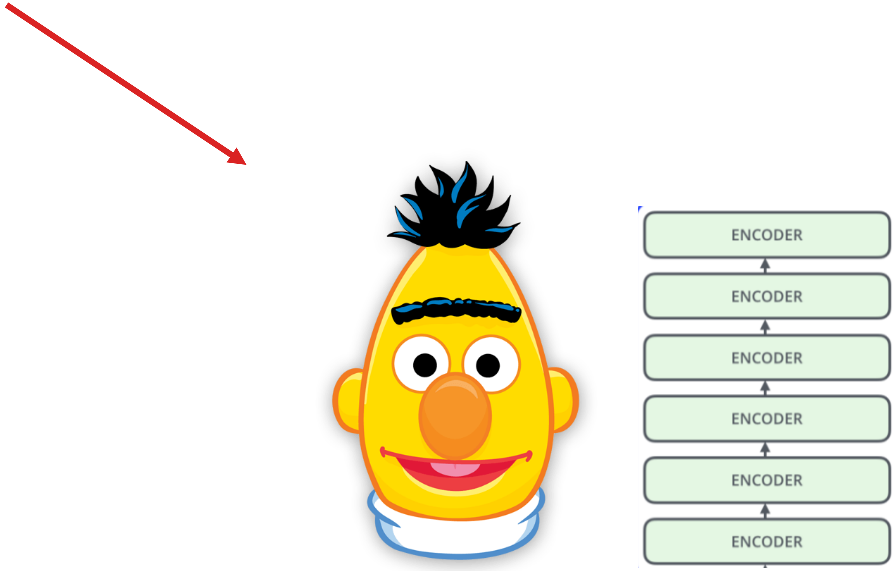


http://jalammar.github.io/illustrated-transformer/

Transformers: Family

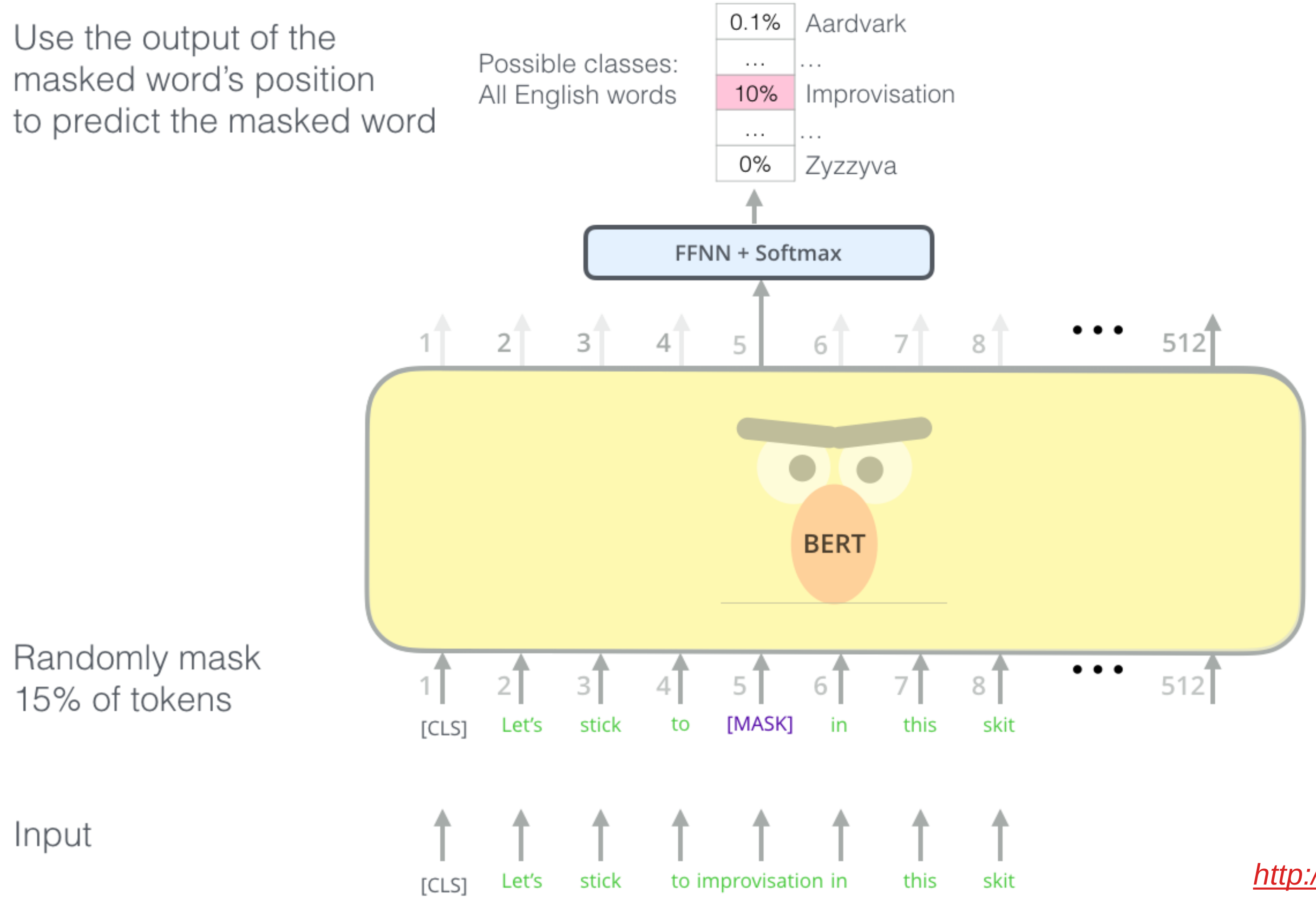






Transfer Learning

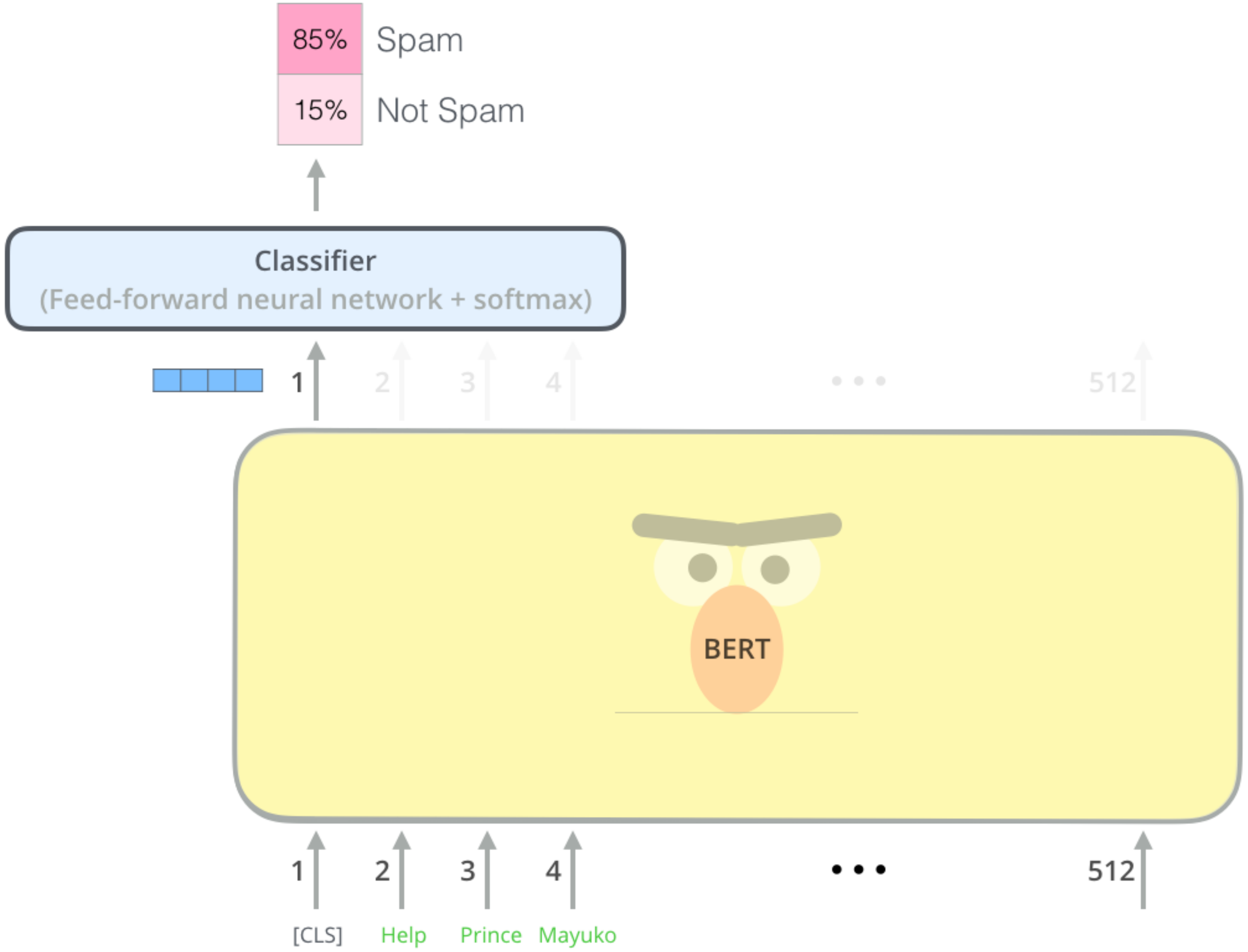
Step 1: Pretraining



http://jalammar.github.io/illustrated-bert/

Transfer Learning

Step 2: Fine-tuning



http://jalammar.github.io/illustrated-bert/

Transformers are extermely versatile:

Multilingual: There are models pretrained on 100 languages!

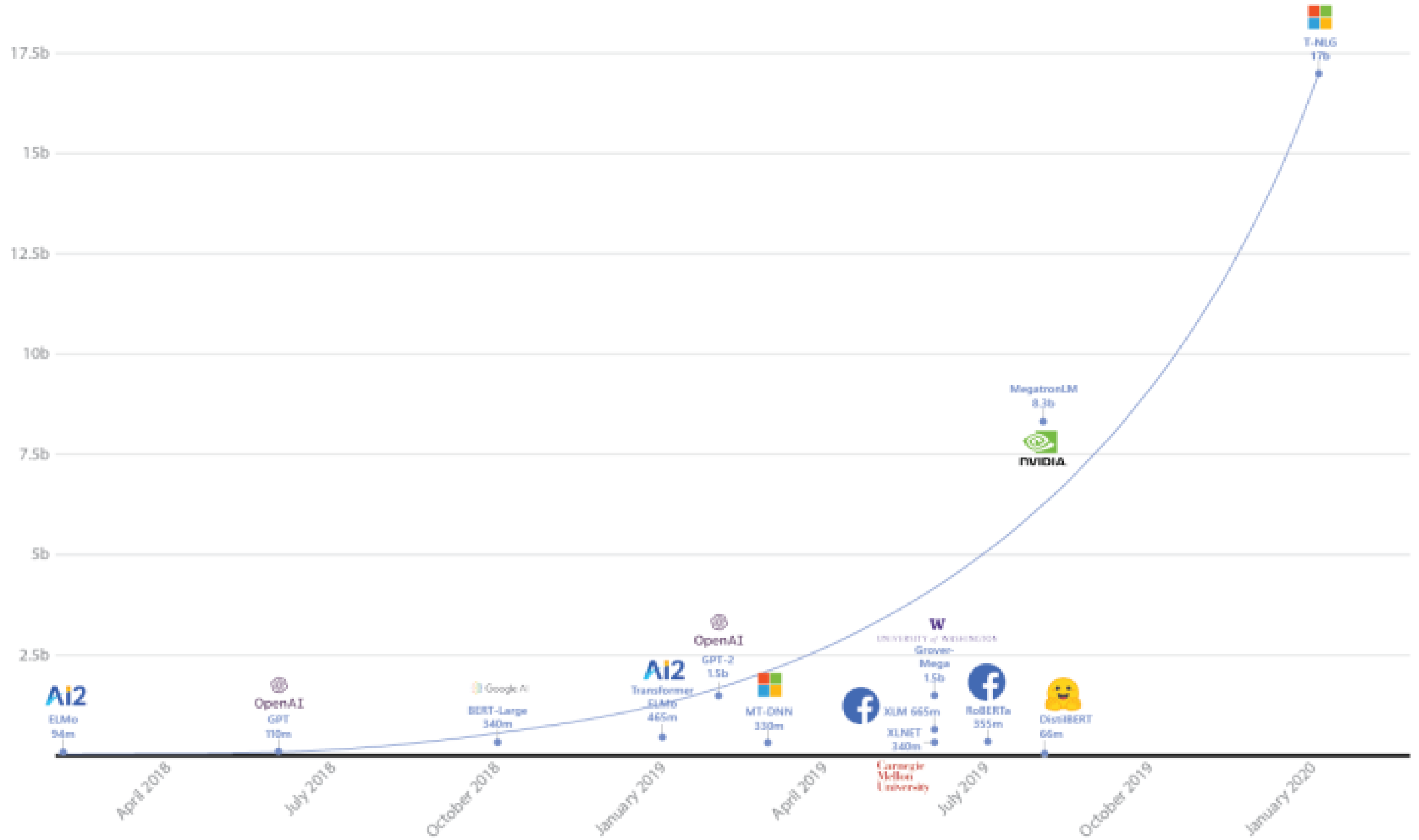
Many tasks:

- Text generation
- Text classification
- Token classification (e.g. NER)
- Question Answering
- Summarization
- Translation

Question Answering with claim texts

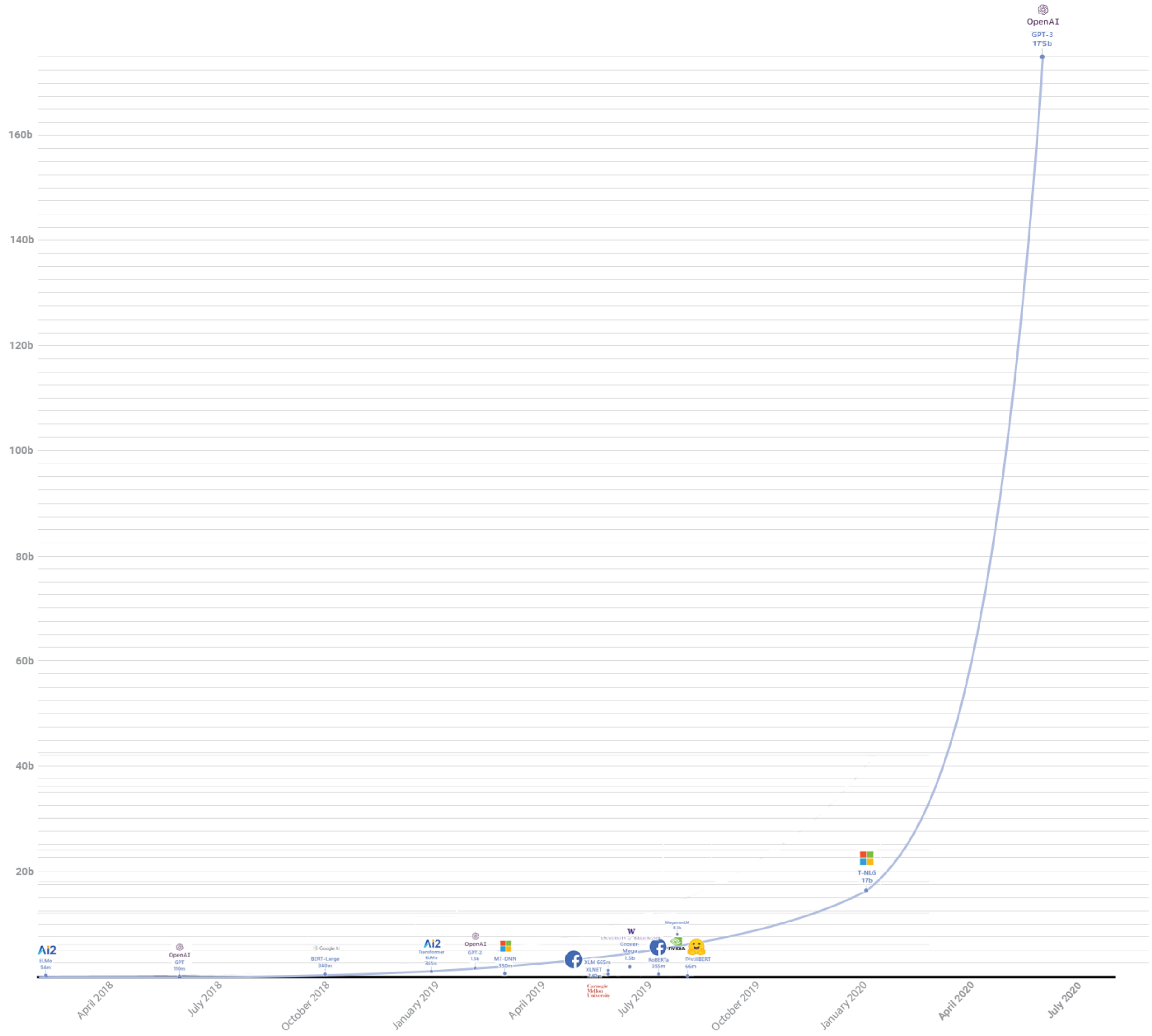
DEMO

Transformers: Scale



https://venturebeat.com/2020/02/10/microsoft-trains-worlds-largest-transformer-language-model/

Transformers: Scale



https://news.t0.vc/XUYH

Transformers: Pros and Cons

Pros

- Contextual embeddings
- Few labels required with transfer learning
- State-of-the-art performance

Cons

- Massive models
- Interpretability/bias
- Finite context

Thank you for your attention!

Mail: leandro.vonwerra1@mobi.ch
Twitter: @lvwerra