AN INTRODUCTION TO NLG

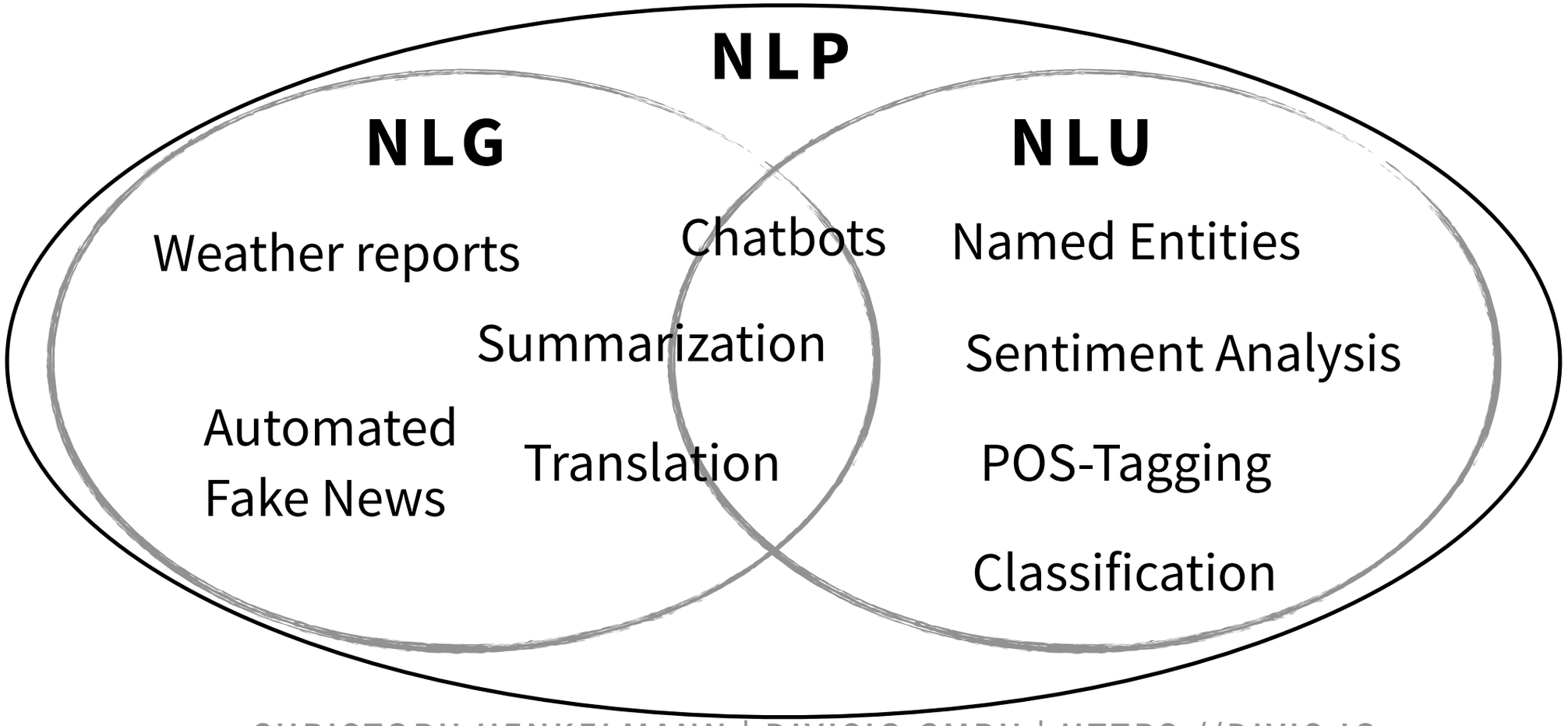
CHRISTOPH HENKELMANN | DIVISIO GMBH | HTTPS://DIVIS.IO

CHRISTOPH HENKELMANN

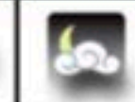
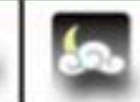
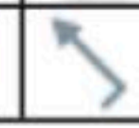
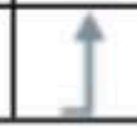
- CTO & Founder DIVISIO
- Studied Computer Science in Bonn (Diplominformatiker)
- 20 years programming experience
- Getting Things Done
- @divisio_ai
- @chenkelmann

NLP, NLG, NLU

- Natural Language = *textual data,*
  not images of text (OCR)
  or spoken language (STT, TTS)
- Natural Language processing (NLP) divides into
  Natural Language Understanding (NLU) and
  Natural Language Generation (NLG)



NLG EXAMPLES: WEATHER REPORTS

9th December, Monday			10th December, Tuesday			11th December, Wednesday			12th December, Thursday		
Morn.	*Aft.*	*Night*	*Morn.*	*Aft.*	*Night*	*Morn.*	*Aft.*	*Night*	*Morn.*	*Aft.*	*Night*
											
											
Min: 1° Max: 14°			Min: 5° Max: 16°			Min: 7° Max: 16°			Min: 11° Max: 15°		

There will be clear skies at the beginning and towards the middle of the term, although at the end they will be very cloudy. We expect precipitations on Thursday morning. The temperatures will be normal for the minimums and high for the maximums for this period of the year, with minimums in notable increase and maximums without changes.

Ramos-Soto et. al.: Linguistic Descriptions for Automatic Generation of Textual Short-Term Weather Forecasts on Real Prediction Data

https://arxiv.org/pdf/1411.4925.pdf

NLG EXAMPLES: FINANCE

Der Börsentag endet für Funkwerk mit einem Minus von -1,75%

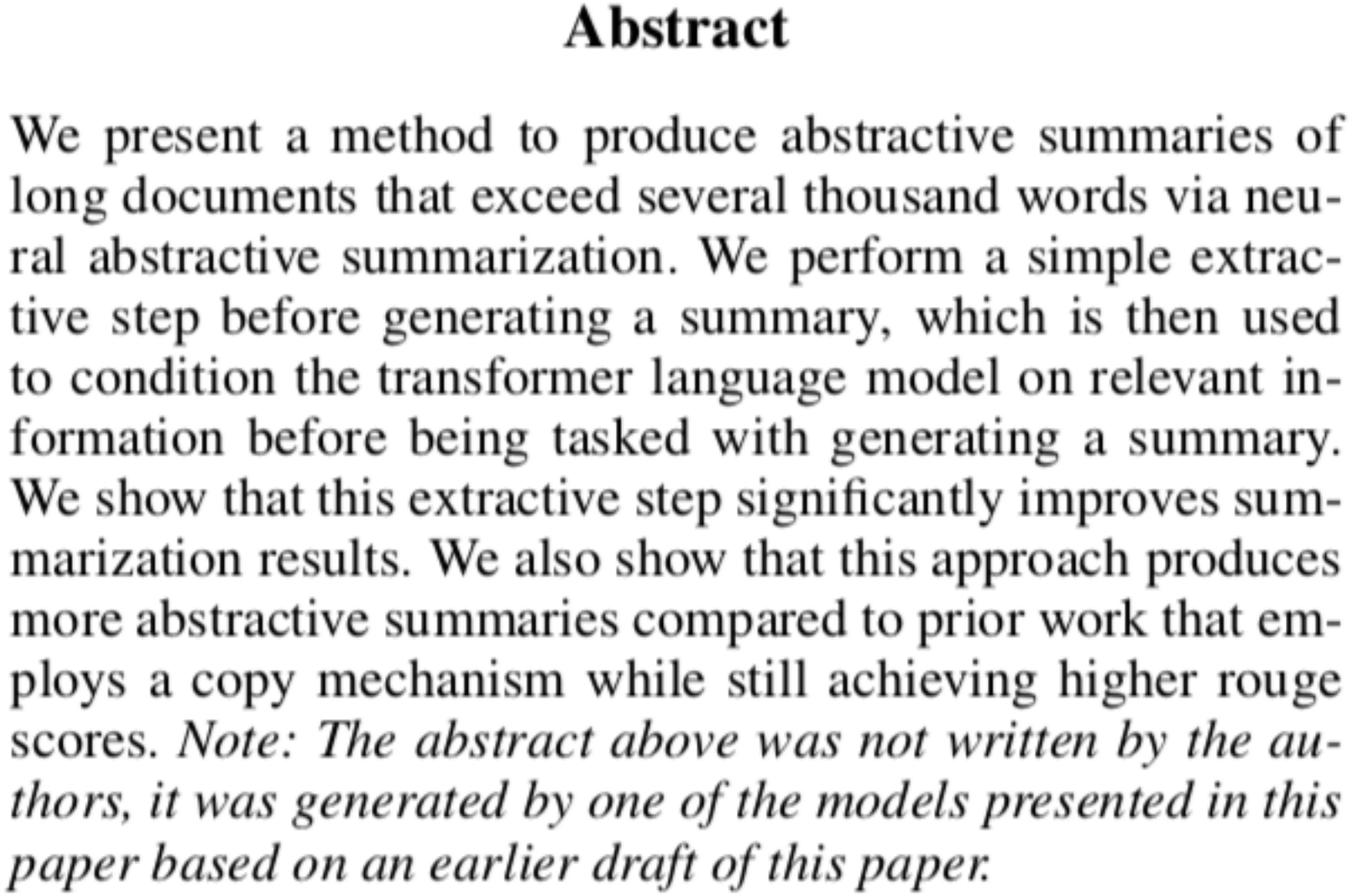
Begleitet von einem überdurchschnittlichen Handelsvolumen ging es für die Funkwerk-Aktie heute nach unten. Bei Eröffnung stand die Aktie noch bei 22,80 EUR. Der Aktienpreis fiel dann im Laufe des Tages um -1,75%, um bei Börsenschluss bei 22,40 EUR zu landen. Intraday notierte die Aktie zwischenzeitlich mit 21,20 EUR augenfällig darunter.

Die Kursentwicklung bei Funkwerk schien in den letzten beiden Börsentagen unentschlossen und pendelt derzeit auf und ab. Dem heutigen Verlust war gestern ein Plus von 0,90% vorangegangen.

Trotz des heutigen Verlustes ergibt sich noch ein bemerkenswerter Gewinn von 19,15% über die letzten zwölf Monate.

DIVISIO Sokratext

NLG EXAMPLES: SUMMARIZATION



Abstract

We present a method to produce abstractive summaries of long documents that exceed several thousand words via neural abstractive summarization. We perform a simple extractive step before generating a summary, which is then used to condition the transformer language model on relevant information before being tasked with generating a summary. We show that this extractive step significantly improves summarization results. We also show that this approach produces more abstractive summaries compared to prior work that employs a copy mechanism while still achieving higher rouge scores. *Note: The abstract above was not written by the authors, it was generated by one of the models presented in this paper based on an earlier draft of this paper.*

Subramanian et. al.: On Extractive and Abstractive Neural Document Summarization with Transformer Language Models

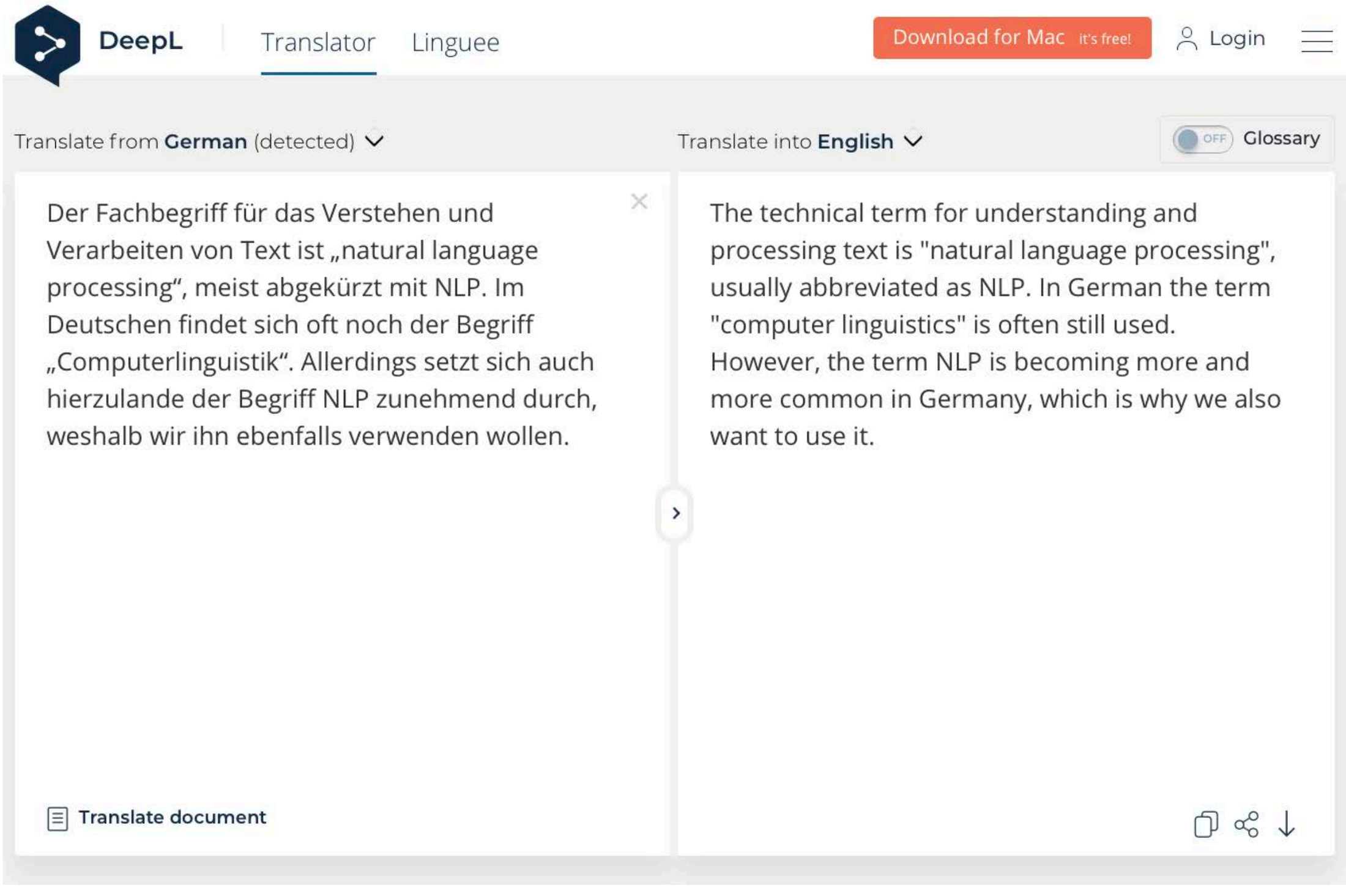
https://arxiv.org/pdf/1909.03186.pdf

NLG EXAMPLES: LANGUAGE MODELLING

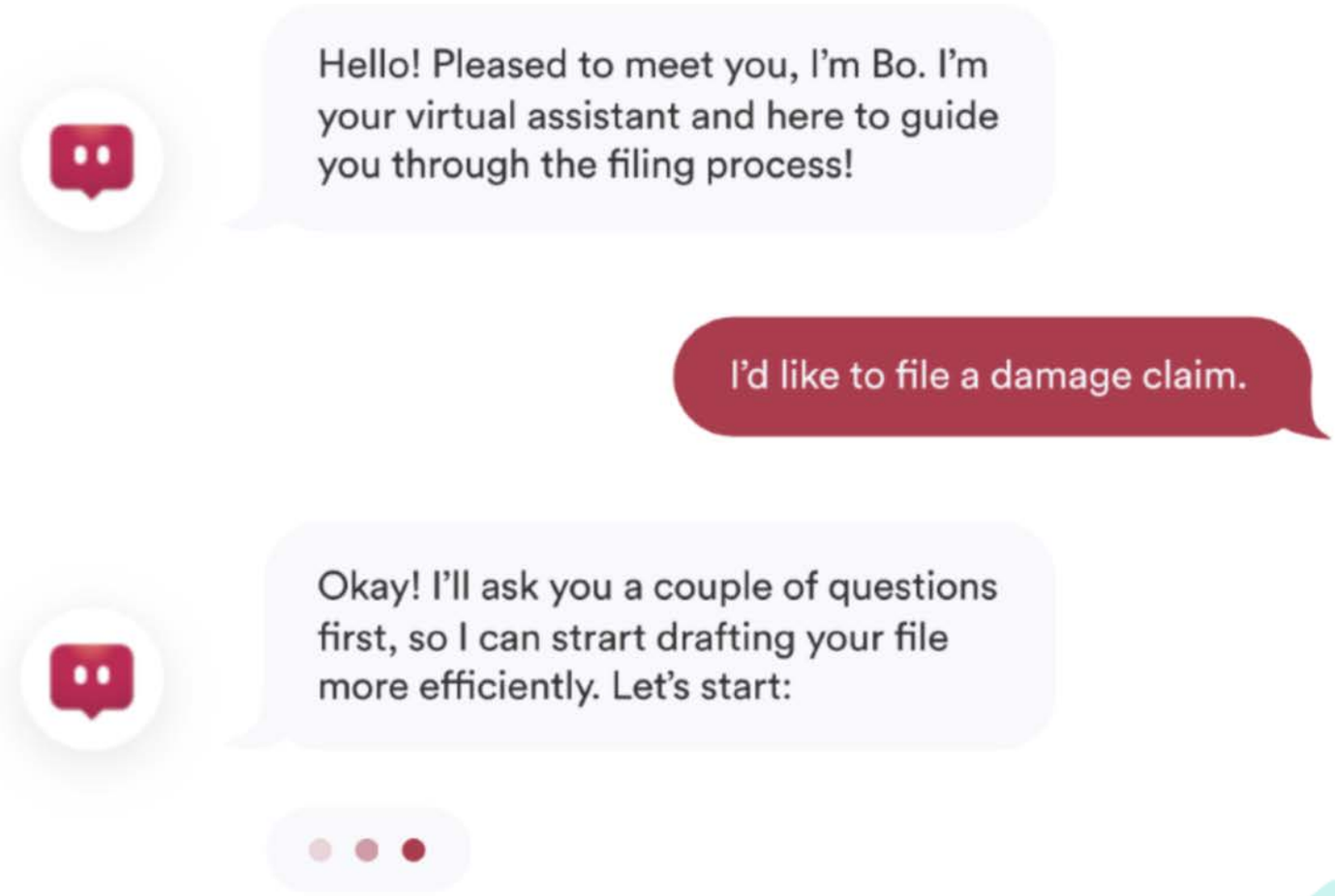
SYSTEM PROMPT (HUMAN-WRITTEN)	*In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.*
MODEL COMPLETION (MACHINE-WRITTEN, 10 TRIES)	The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science. Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved. Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow. Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez. Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them – they were so close they could touch their horns. While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, "We can see, for example, that they have a common 'language,' something like a dialect or dialectic."

https://openai.com/blog/better-language-models/

NLG EXAMPLES: TRANSLATION



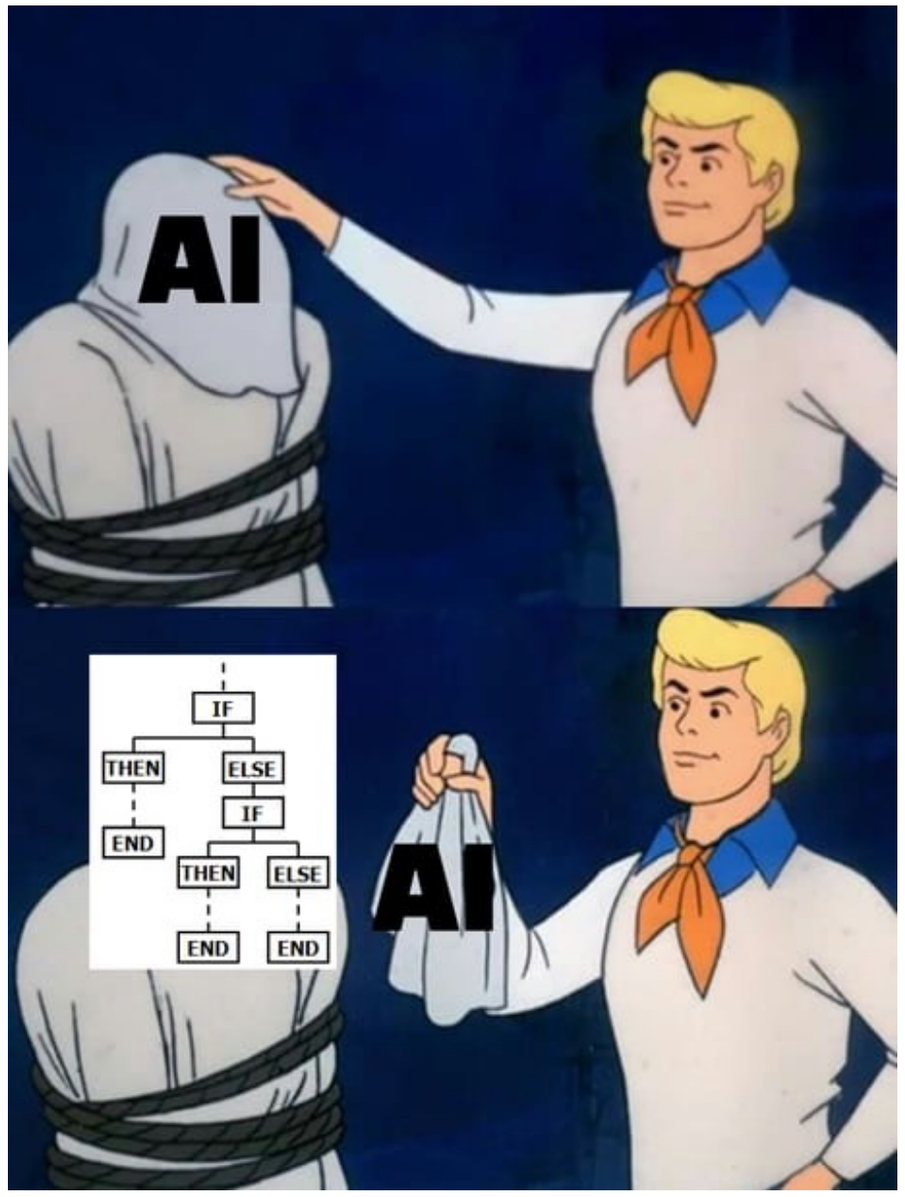
NLG EXAMPLES: CHATBOTS



Chatlayer.ai

THE DIRTY TRUTH ABOUT COMMERCIAL (DATA DRIVEN)NLG

Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation, Gatt et. al., 29.01.2018
https://arxiv.org/pdf/1703.09902.pdf



- Rules, hacks, templates, coding: still SOTA for controllable content
- Deep Learning / Stochastic models: Little/imprecise control over result (except translation)
- ⇒ Smart template engines, “text robots”, etc. are still the norm

- But: Things are changing fast! (GPT-3 etc.)

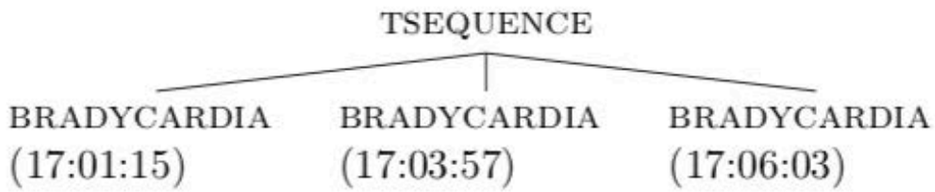
NLG WITH "GOOD OLD AI": SUBTASKS

- *Content determination*: Deciding which information to include in the text under construction
- *Text structuring*: Determining in which order information will be presented in the text
- *Sentence aggregation*: Deciding which information to present in individual sentences
- *Lexicalisation*: Finding the right words and phrases to express information
- *Referring expression generation*: Selecting the words and phrases to identify domain objects
- *Linguistic realisation*: Combining all words and phrases into well-formed sentences.

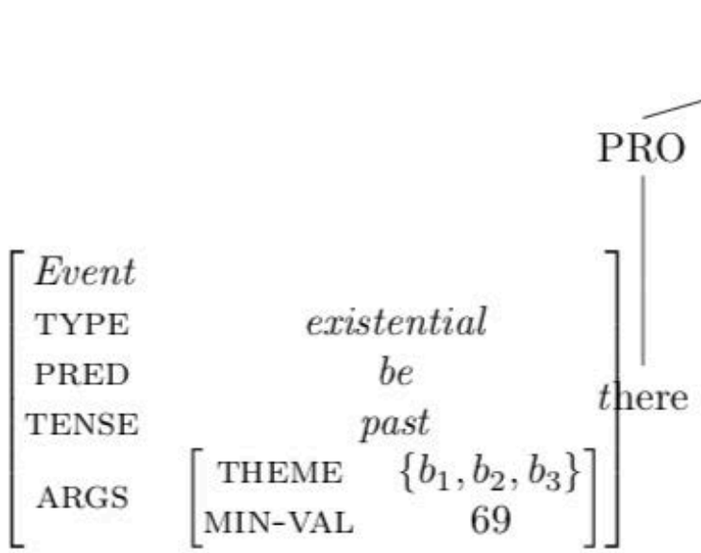
Gatt et. al.: Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation
https://arxiv.org/pdf/1703.09902.pdf

NLG WITH "GOOD OLD AI": SUBTASKS

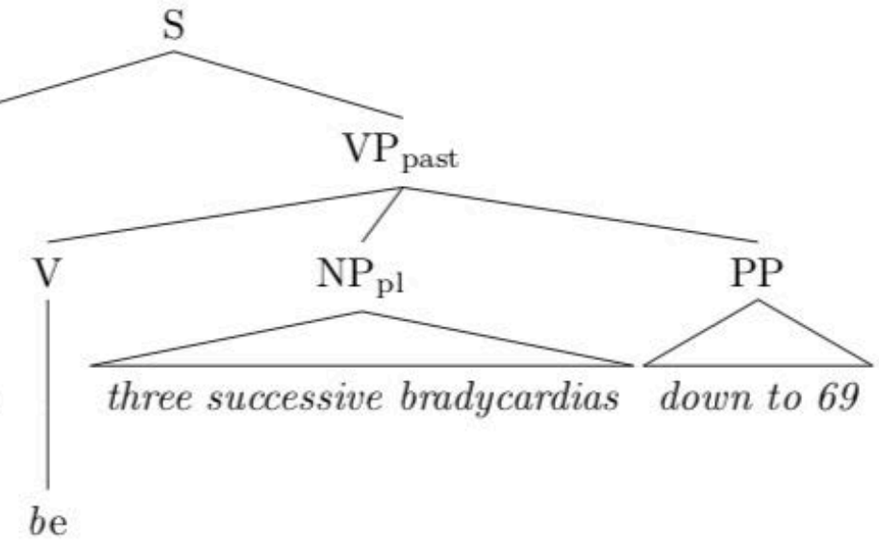
(a) Content Determination



(b) Text Structuring



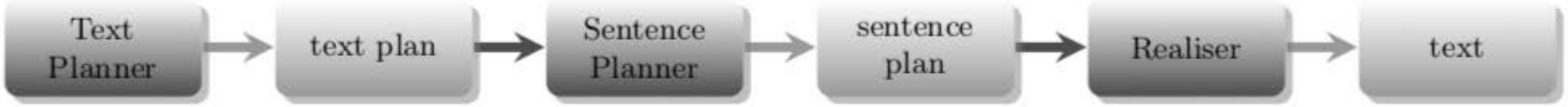
(c) Lexicalisation etc.



(d) Realisation

Gatt et. al.: Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation
https://arxiv.org/pdf/1703.09902.pdf

NLG WITH "GOOD OLD AI": PIPELINE



Gatt et. al.: Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation
https://arxiv.org/pdf/1703.09902.pdf

NLG WITH “GOOD OLD AI”: ADVANTAGES

- Precise control over content (What)
- Precise control over style (How)
- Results are explainable (Why)

Photo by Benjamin Rascoe on Unsplash

NLG WITH “GOOD OLD AI”: DISADVANTAGES

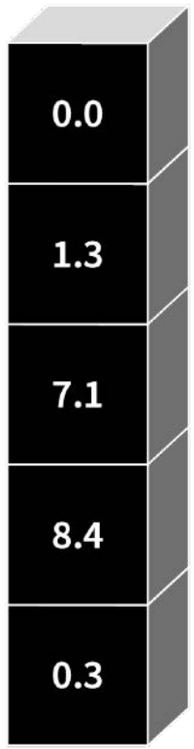
- Repetitive
- Uncreative
- Very narrow focus
- Clumsy formulations
- A lot of manual work

DEEP LEARNING BASED NLG

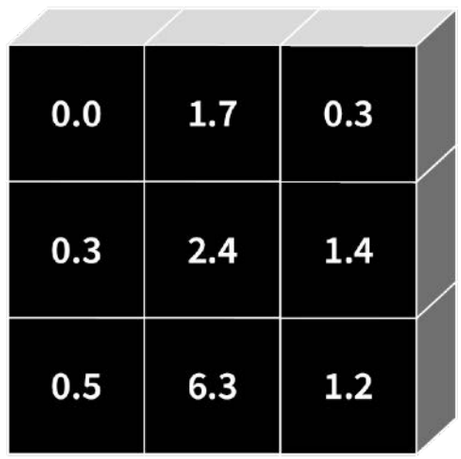
NO, you cannot understand the meaning of a text without explicitly evaluating its linguistic constituents and defining grammar rules!

haha gpus go brrrrrrrrr

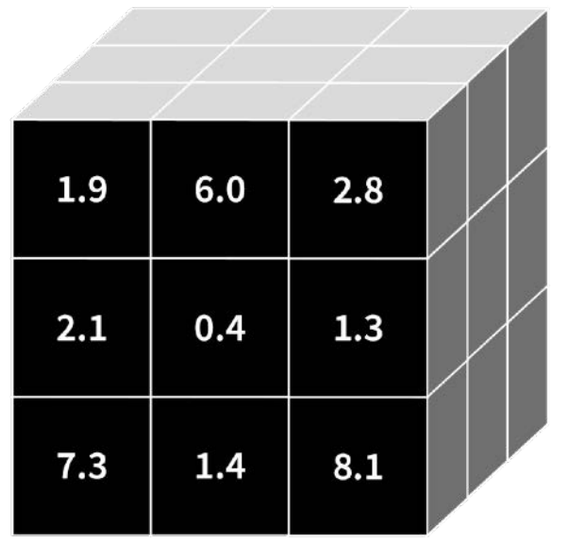
EXPRESSING WORDS AS TENSORS



1-D Tensor
(e.g. a column in a spreadsheet)

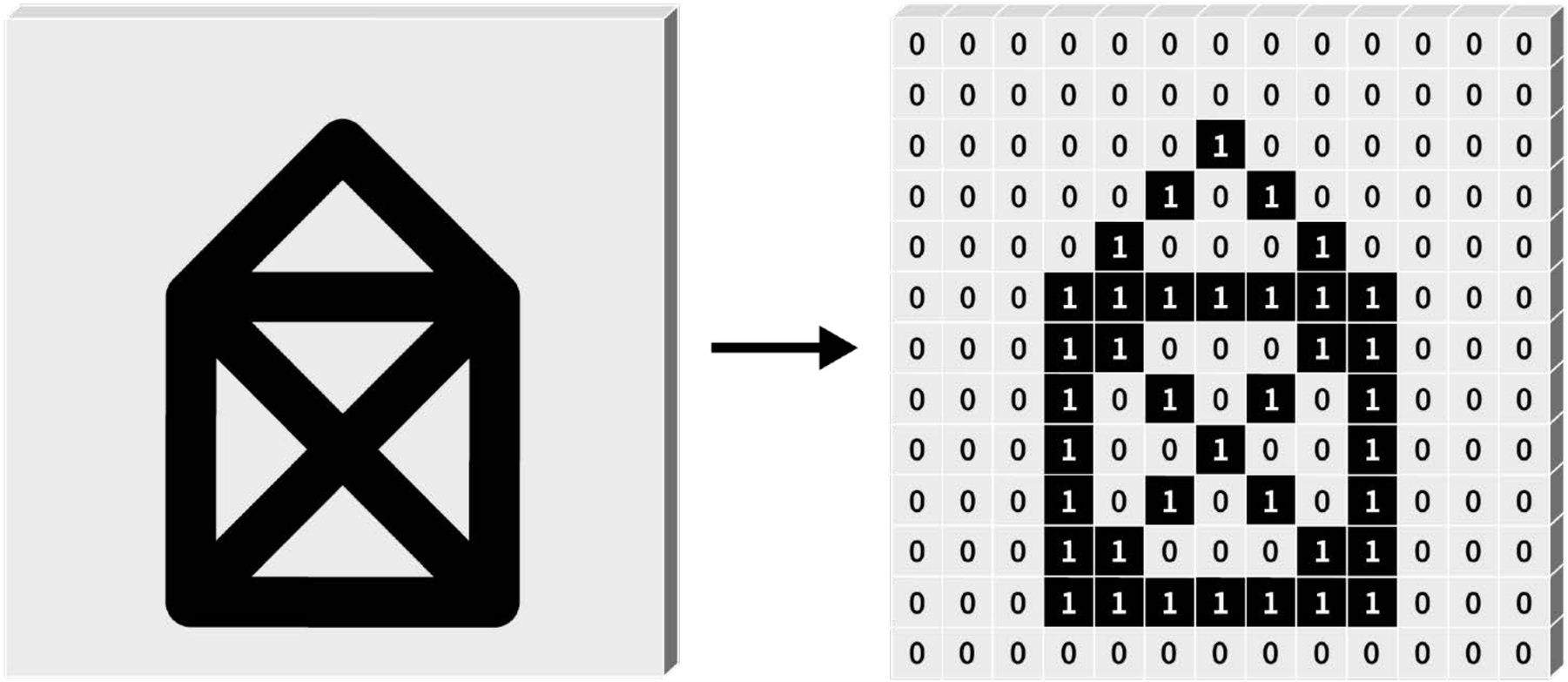


2-D Tensor
(e.g. for black-and-white images)



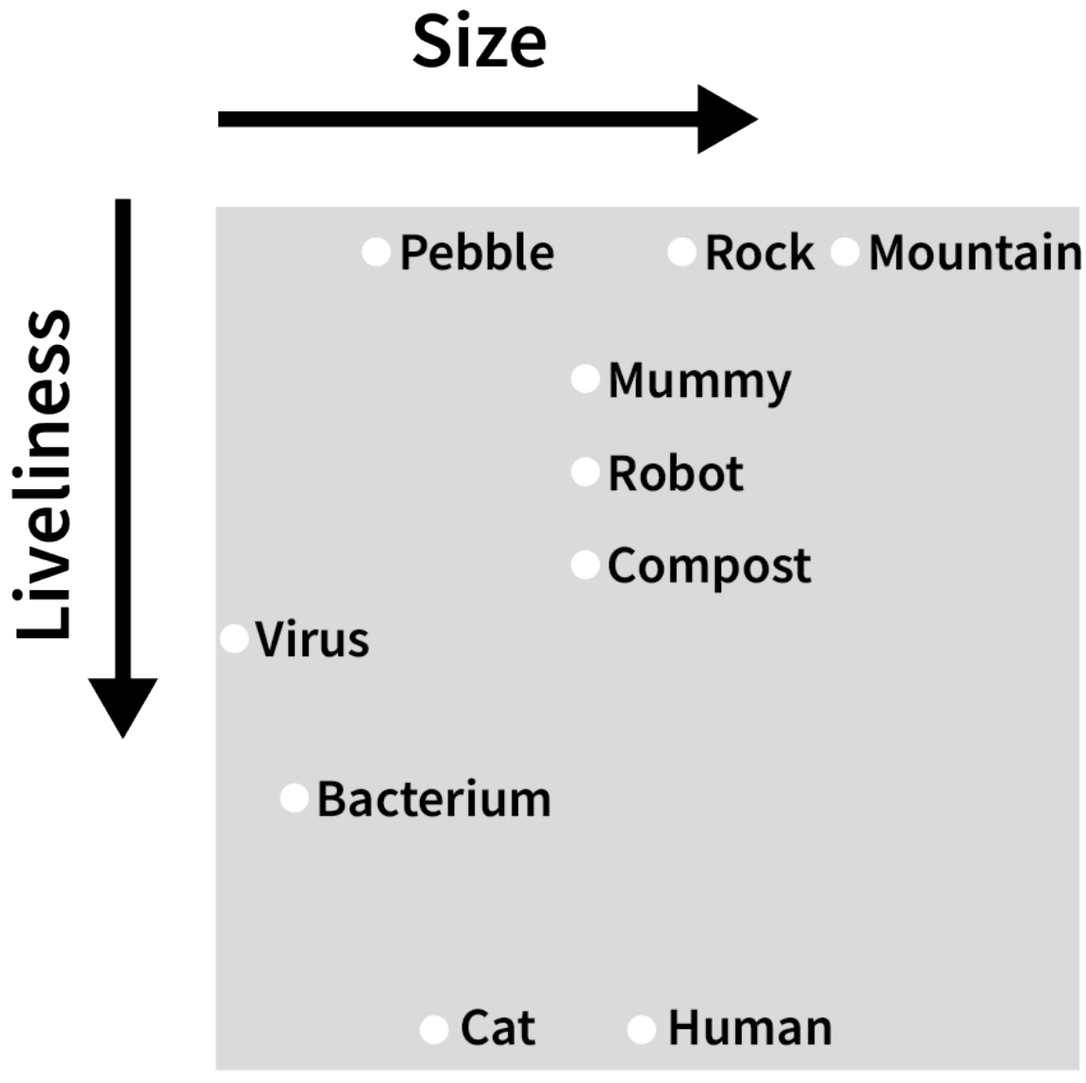
3-D Tensor
(e.g. for color images, every layer for one color channel)

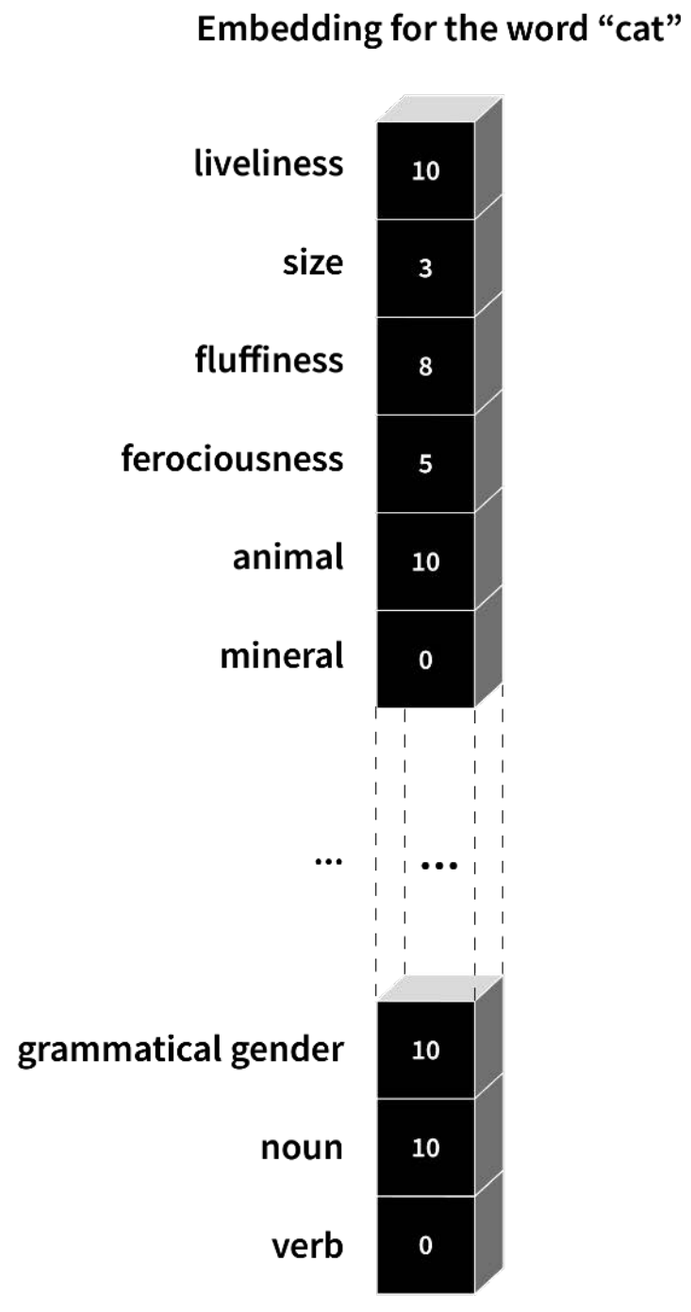
EXPRESSING WORDS AS TENSORS



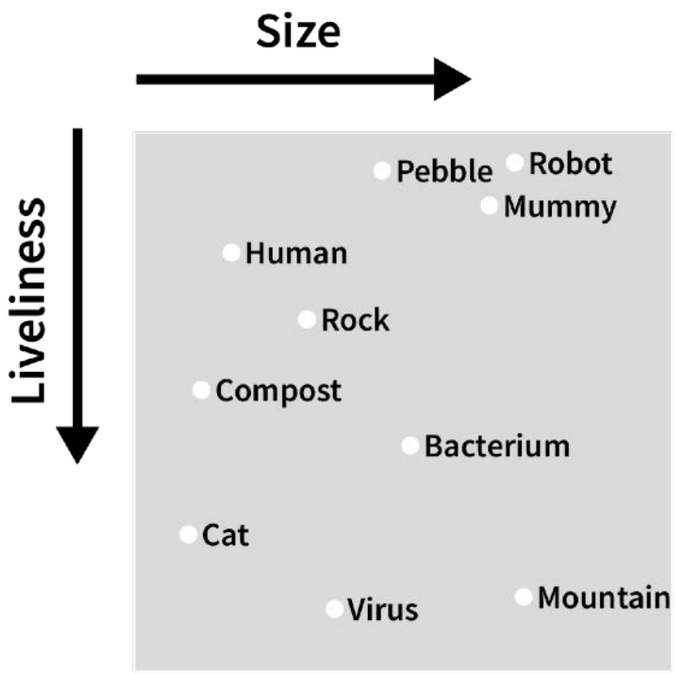
“house” → ?

WORD (TOKEN) EMBEDDINGS

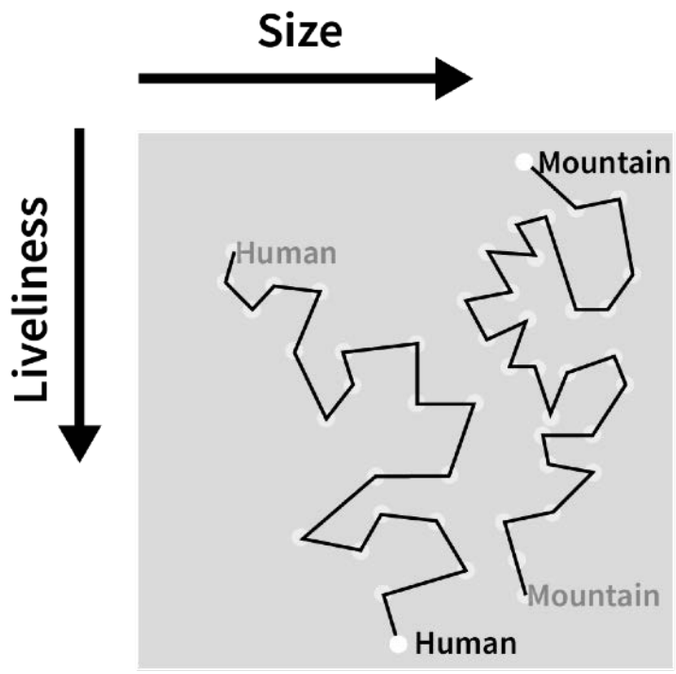




LEARNING EMBEDDINGS

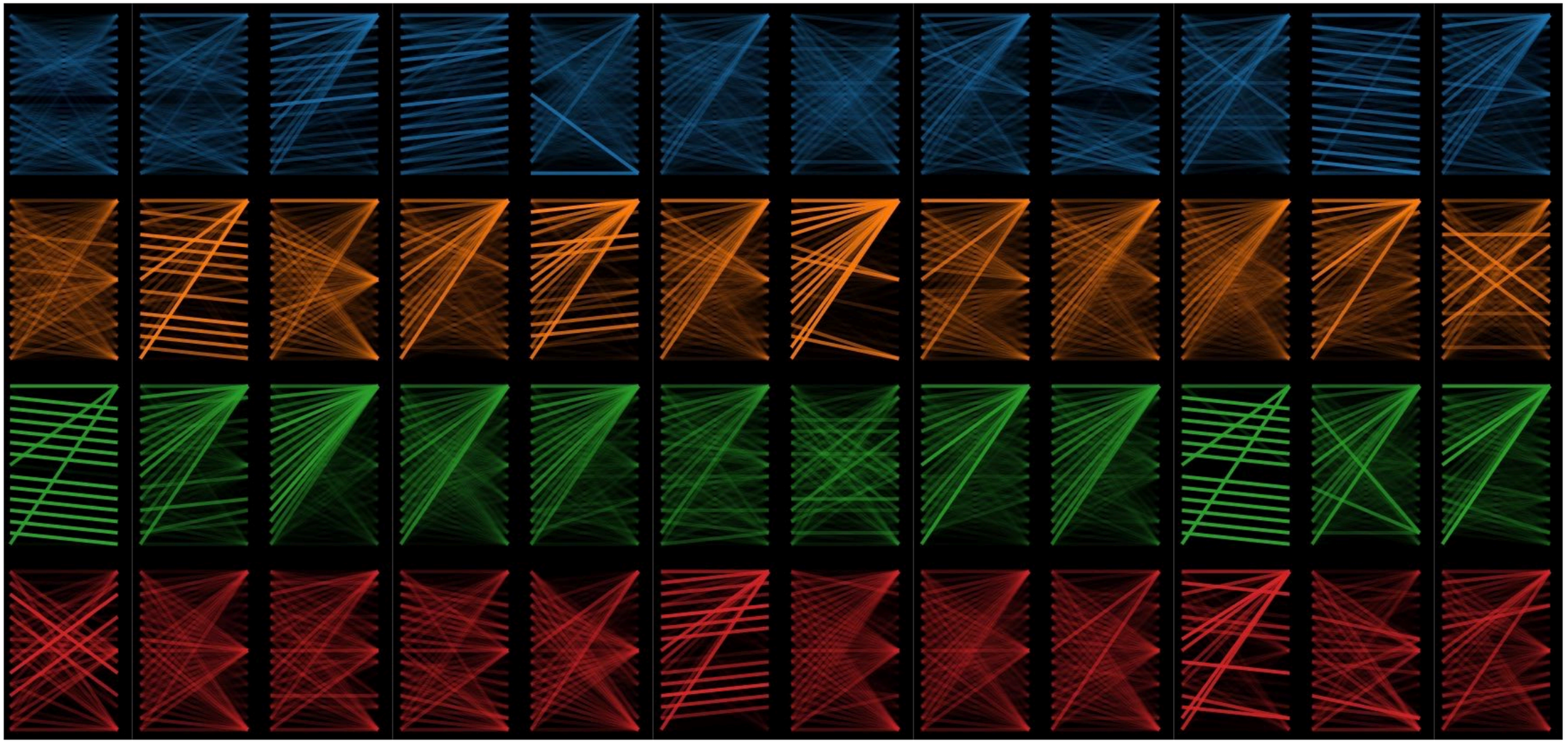


Start of training:
the positions are
completely random



During training
the axis positions are
adjusted step by step

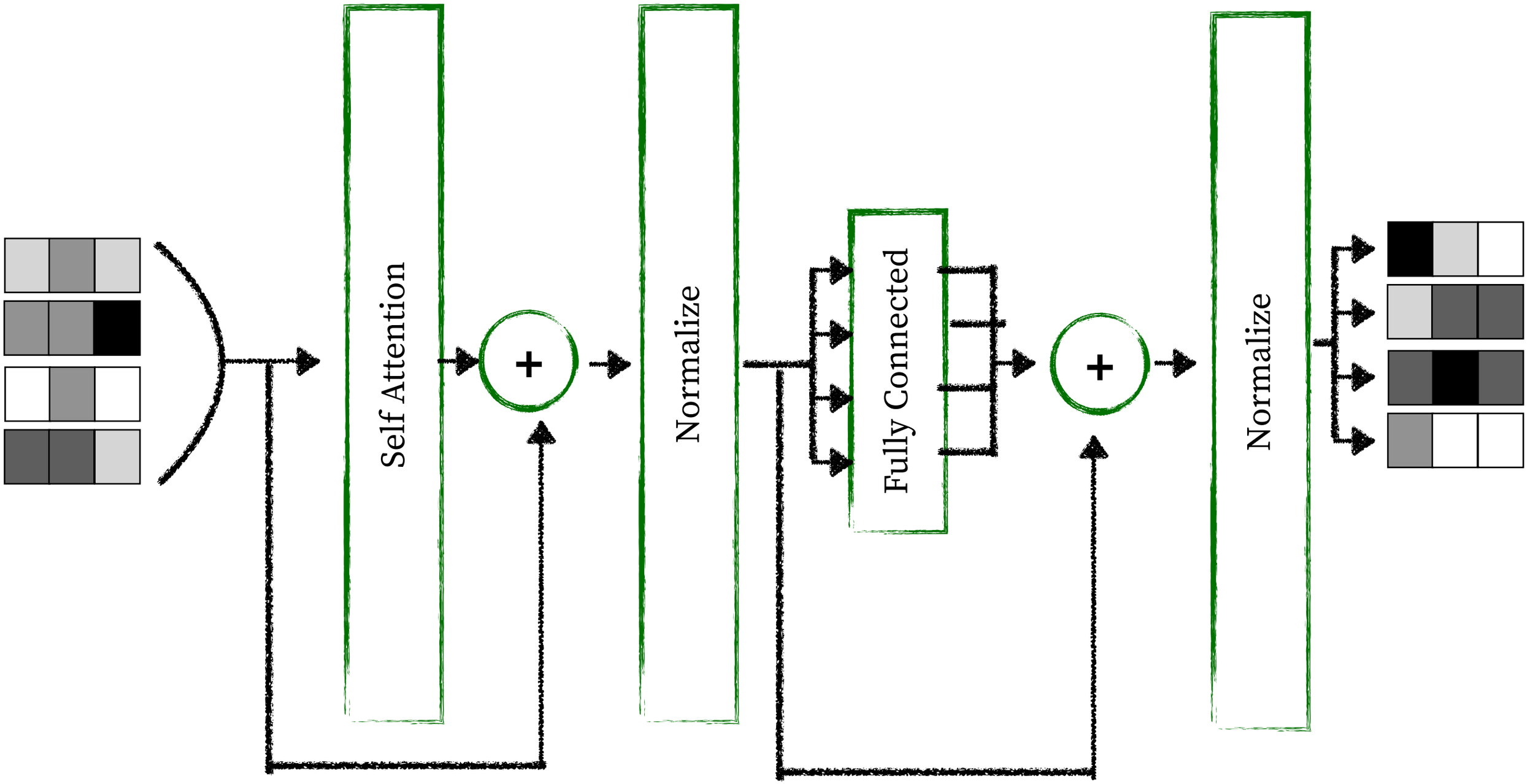
ATTENTION LAYERS



https://github.com/jessevig/bertviz

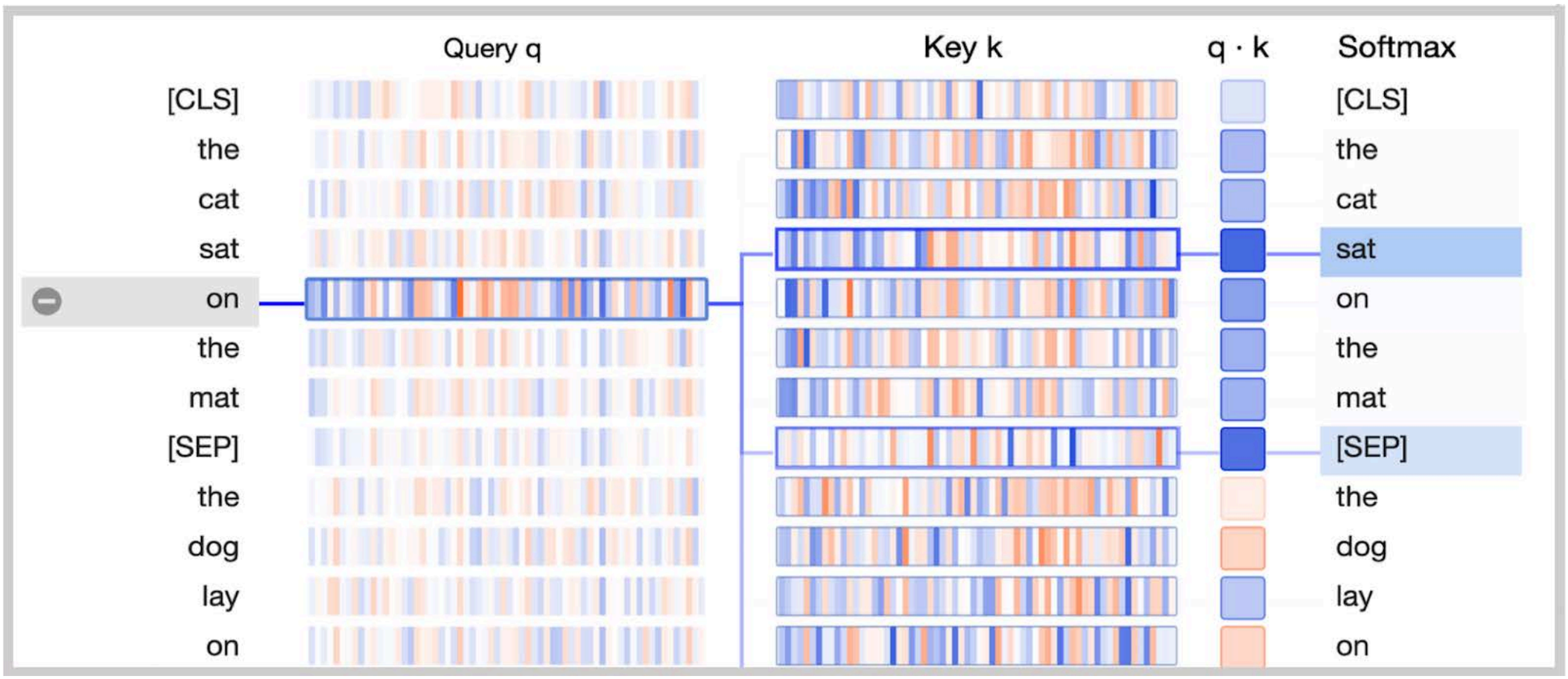
BERT STRUCTURE - TRANSFORMER BLOCK

Attention is all you need, Vaswani et. al., 06.12.2017



(SELF-)ATTENTION LAYERS

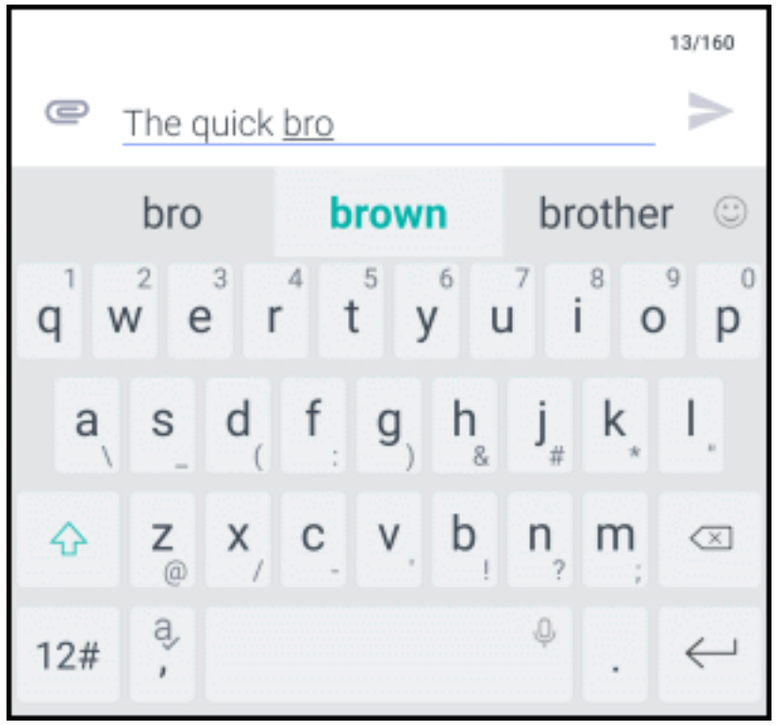
- Embeddings are run through learned linear projections to create keys & queries for each token
- Using the similarity of q & k, each embedding is replaced by a weighted sum of the other embeddings



https://github.com/jessevig/bertviz

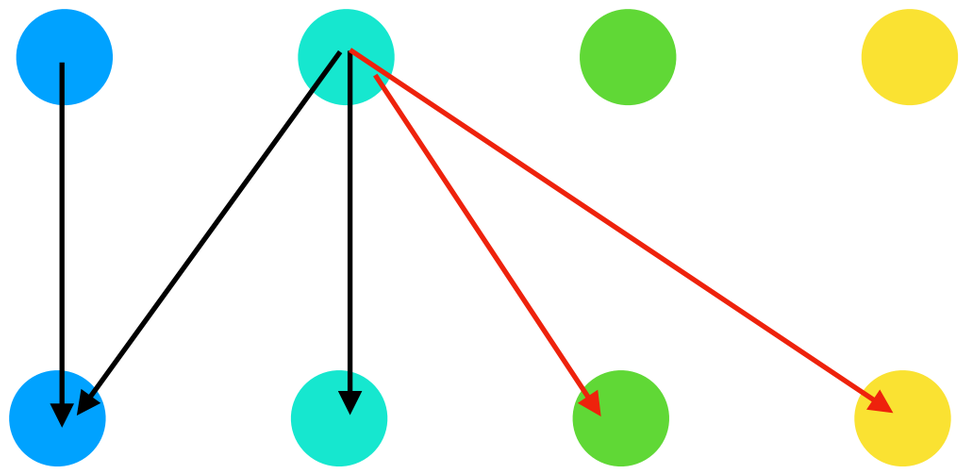
LANGUAGE MODELLING

- Language Modelling (LM) is the task of predicting the next words / tokens given an input
- Can be used to create huge amounts of training data from unlabelled text

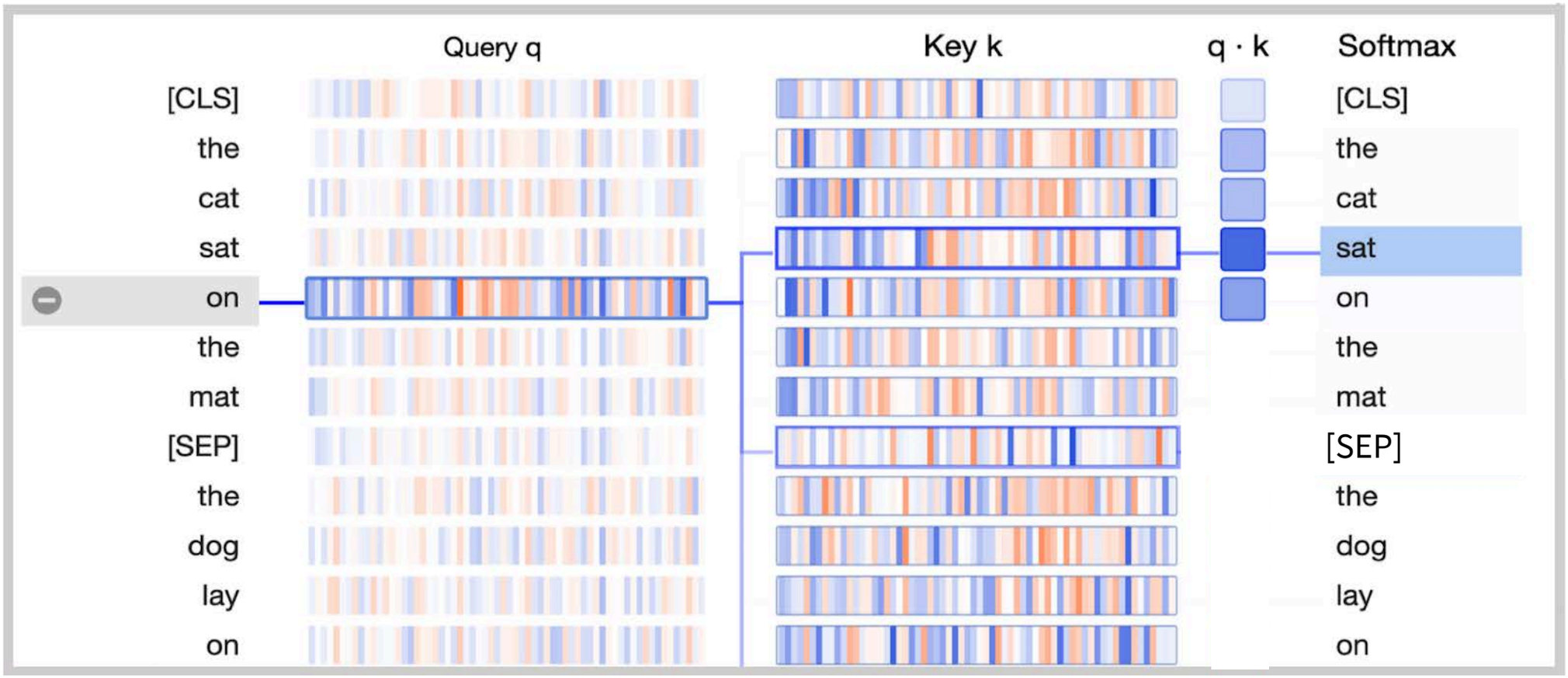


LANGUAGE MODELLING WITH TRANSFORMERS

- Predict every word at once
- To prevent cheating, we need “masked” attention
- Every word can only “see” its predecessors
- “Autoregressive Transformer Models”



MASKED (SELF-)ATTENTION LAYERS



https://github.com/jessevig/bertviz

GPT ARCHITECTURE & TRAINING

- Input is tokenized with Byte Pair Coding
  ("Carbon dioxide"=>['_C', 'ar', 'b', 'on', '_d', 'i', 'o', 'x', 'ide'])
- Each token is mapped to an input embedding
- Embeddings are initialized randomly (like weights)
- The net is a huge stack of masked transformers
- GPT is trained only on the LM task: it reproduces its input

WHAT IS THE DIFFERENCE BETWEEN GPT-1, -2 & -3?

- GPT-2 is bigger than GPT-1
- GPT-3 is bigger than GPT-2

HOW TO CREATE TEXT WITH GPT?

- Give GPT a prompt
- Let GPT finish it token by token (using beam search)

INPUT TO "PRIME" GPT:

```
Q: What is human life expectancy in the United States?
A: Human life expectancy in the United States is 78 years.

Q: Who was president of the United States in 1955?
A: Dwight D. Eisenhower was president of the United States in 1955.

Q: What party did he belong to?
A: He belonged to the Republican Party.

Q: Who was president of the United States before George W. Bush?
A: Bill Clinton was president of the United States before George W. Bush.

Q: Who won the World Series in 1995?
A: The Atlanta Braves won the World Series in 1995.
```

examples by Kevin Lacker (https://lacker.io/ai/2020/07/06/giving-gpt-3-a-turing-test.html)

HOW TO CREATE TEXT WITH GPT?

AFTER THE INPUT, ADD ANOTHER QUESTION, LET GPT GENERATE THE ANSWER

```
Q: What is your favorite animal?
A: My favorite animal is a dog.

Q: Why?
A: Because dogs are loyal and friendly.

Q: What are two reasons that a dog might be in a bad mood?
A: Two reasons that a dog might be in a bad mood are if it is hungry or if it is hot.

[...]

Q: How many eyes does a spider have?
A: A spider has eight eyes.

[...]

Q: How many eyes does a blade of grass have?
A: A blade of grass has one eye.

Q: How many rainbows does it take to jump from Hawaii to seventeen?
A: It takes two rainbows to jump from Hawaii to seventeen.

Q: Which colorless green ideas sleep furiously?
A: Ideas that are colorless, green, and sleep furiously are the ideas of a sleep furiously.

Q: Do you understand these questions?
A: I understand these questions.
```

examples by Kevin Lacker (https://lacker.io/ai/2020/07/06/giving-gpt-3-a-turing-test.html)

HOW TO CREATE TEXT WITH GPT?

- GPT can be steered in a certain direction
- GPT cannot be forced to use certain information
- GPT will fill out the gaps with whatever it sees fit

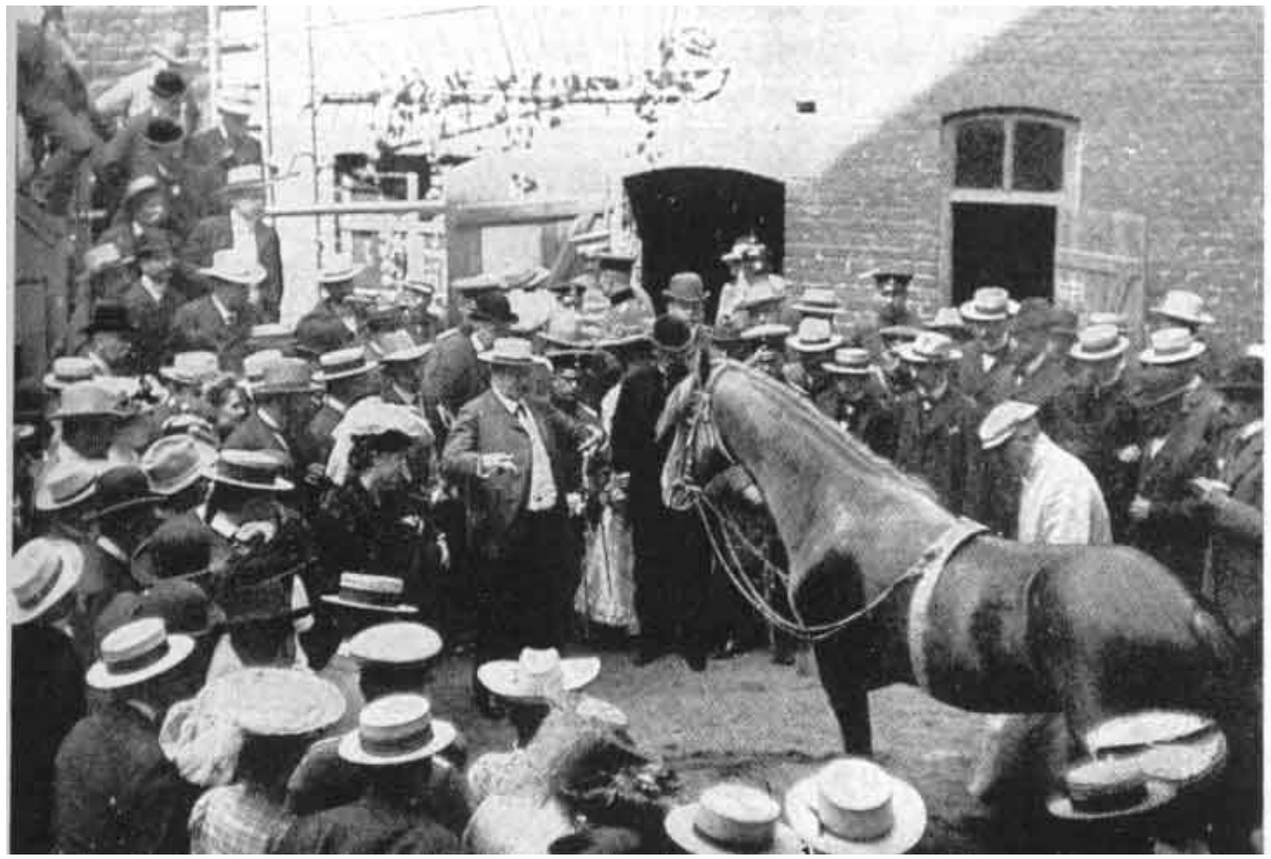
NLG WITH DEEP LEARNING: PROS

- Extremely realistic output
- Varied, creative output
- Hard to tell apart from human output
- Once trained, much less work to adopt to new domain
- Reusable

Photo by MILKOVÍ on Unsplash

NLG WITH DEEP LEARNING: CONS

- Clever Hans models
- No precise control over content
- No precise control over wording
- Makes up facts



HYBRID MODELS

- Replace parts of the “classic” NLG pipeline with DL solutions
- Used at DIVISIO to keep control over content and limit possible formulations
- Adds flexibility to rule-based approach
- Saves time
- Still much more work than fully DL solution, but retains control

COMBINING RULES AND DL

- *Content determination*: Deciding which information to include in the text under construction
- *Referring expression generation*: Selecting the words and phrases to identify domain objects
- *Lexicalisation*: Finding the right words and phrases to express information
- *Text structuring*: Determining in which order information will be presented in the text
- *Sentence aggregation*: Deciding which information to present in individual sentences
- *Linguistic realisation*: Combining all words and phrases into well-formed sentences.

THANK YOU

@divisio_ai

CHRISTOPH HENKELMANN | DIVISIO GMBH | HTTPS://DIVIS.IO