Sentiment Analysis with BERT

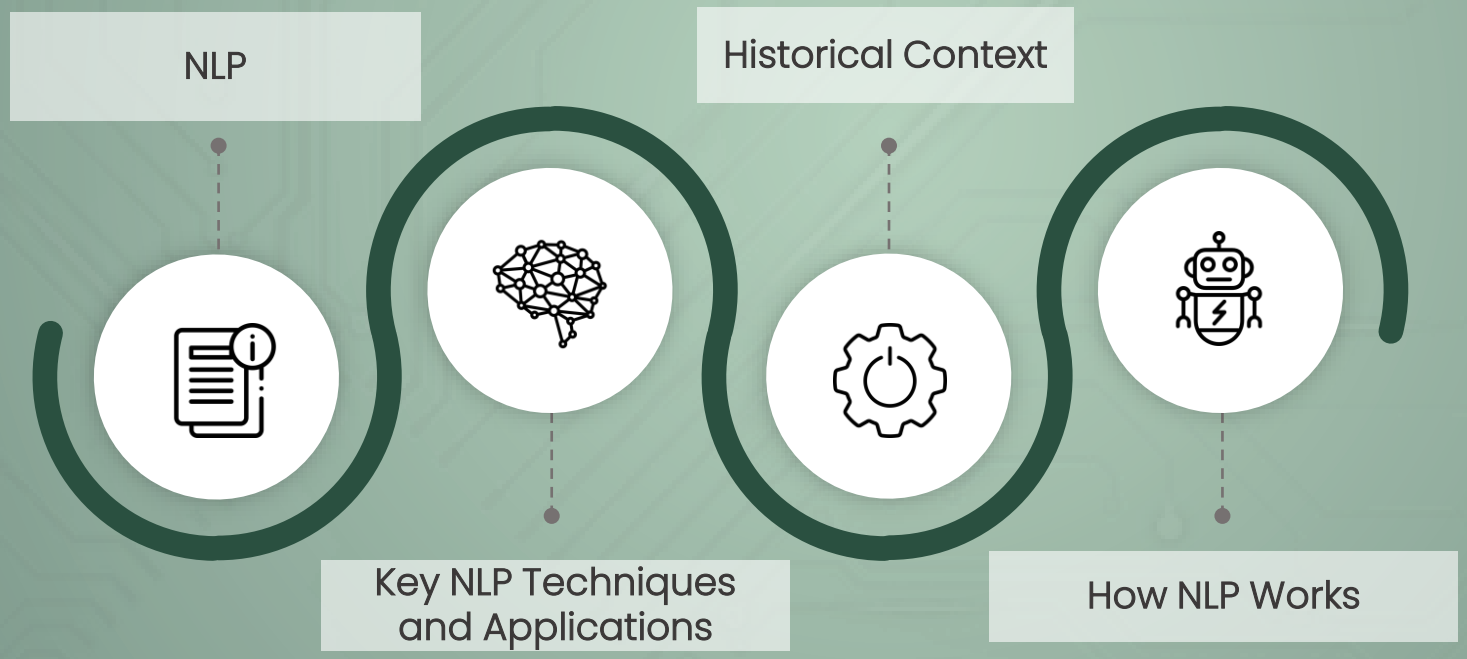
Advanced Software-Engineering
Dr. Harald Stein, Prof. Dr.-Ing. Stefan Edlich
Feb 2024

Agenda

- Understanding Natural Language Processing
- Vector Representations
- Using Transformers for Sentiment Analysis
- Programming Tools
- Example: Sentiments of movie comments

Natural Language Processing

... is a technology that bridges the communication gap between human language and computer understanding.



Natural Language Processing (NLP)

The primary goal of NLP is to enable computers to understand, interpret, and generate human language in a way that is both meaningful and useful.

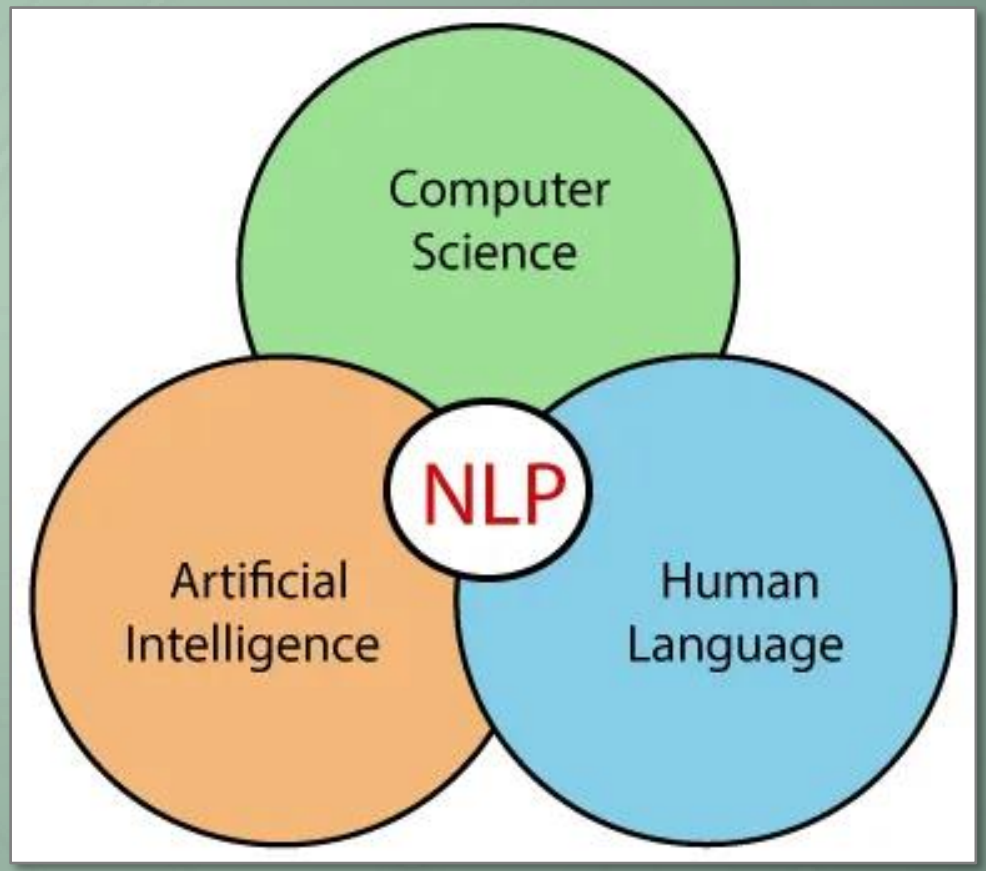
Interdisciplinary Nature

It encompasses areas of

- computer science
- artificial intelligence
- linguistics
- to interpret, recognize, and generate human language in a way that is valuable.

Real-World Applications

- Voice assistants
- Chatbots
- Translation services
- Sentiment analysis
- Customer service.



Key NLP Techniques and Applications

Diverse Applications of NLP

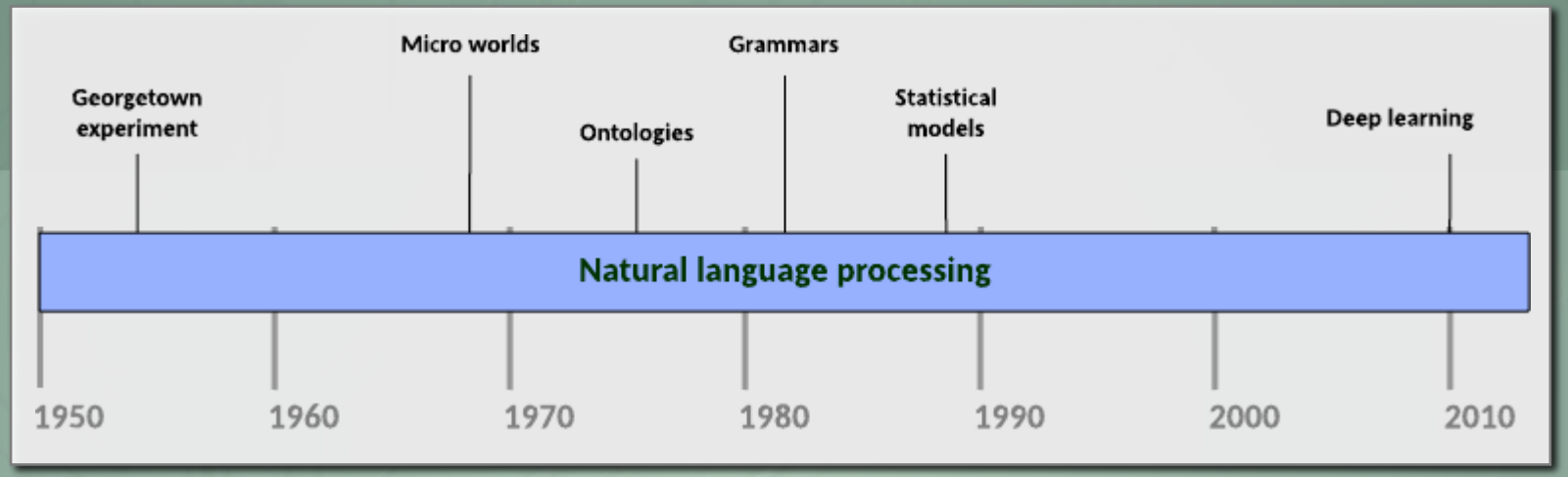
Lecture focus	Description	Kind of task
Sentiment Analysis	Identifying emotions in text to gauge sentiments like positive, negative, or neutral.	Classification
Text/Document Classification:	Assigning categories to text based on content. Utilizes supervised learning on labeled data.	Classification
Part-of-speech (POS) Tagging	Assigning grammatical categories to words in sentences to identify their syntactic roles.	Classification
Language Detection & Machine Translation	Identifying a text's language and translating text between languages	Translation
Information Retrieval	Retrieving relevant information from vast text datasets in response to user queries.	Text Generation

Key NLP Techniques and Applications

Diverse Applications of NLP

	Description	Kind of task
Text Summarization	Condensing long texts while preserving key information and context.	Text Generation
Knowledge Graph & QA System	Organizing information in a structured form and answering questions using that knowledge.	Text Generation
Topic Modeling	Uncovering hidden topics in text collections using unsupervised learning.	(Unsupervised) latent class identification
Text Generation	Automatically generating human-like text.	Text Generation
Speech to Text	Converting spoken language into written text.	Text Generation

Historical context



1950s-1960s: Early Days

- Initial experiments in machine translation and automated reasoning.
- Example: Georgetown-IBM experiment, 1954.Real-World Applications

1980-2010: Statistical Revolution

- Statistical models, algorithms like Hidden Markov Models.
- Machine Learning, Language Processing

2010s: Deep Learning Breakthroughs

- Adoption of deep learning and neural networks.
- Emergence of models like Word2Vec and BERT.

2020s: Advanced Language Models

- State-of-the-art models like GPT and Transformer architectures.
- Unprecedented capabilities in language generation and understanding.

How NLP Works

Generic workflow of Natural Language Processing

Input:
Receives text or speech.

Preprocessing:
Cleans and converts input. Includes tokenization and stemming.

Context Analysis:
Understands structure and meaning. Uses parsing and semantic analysis.

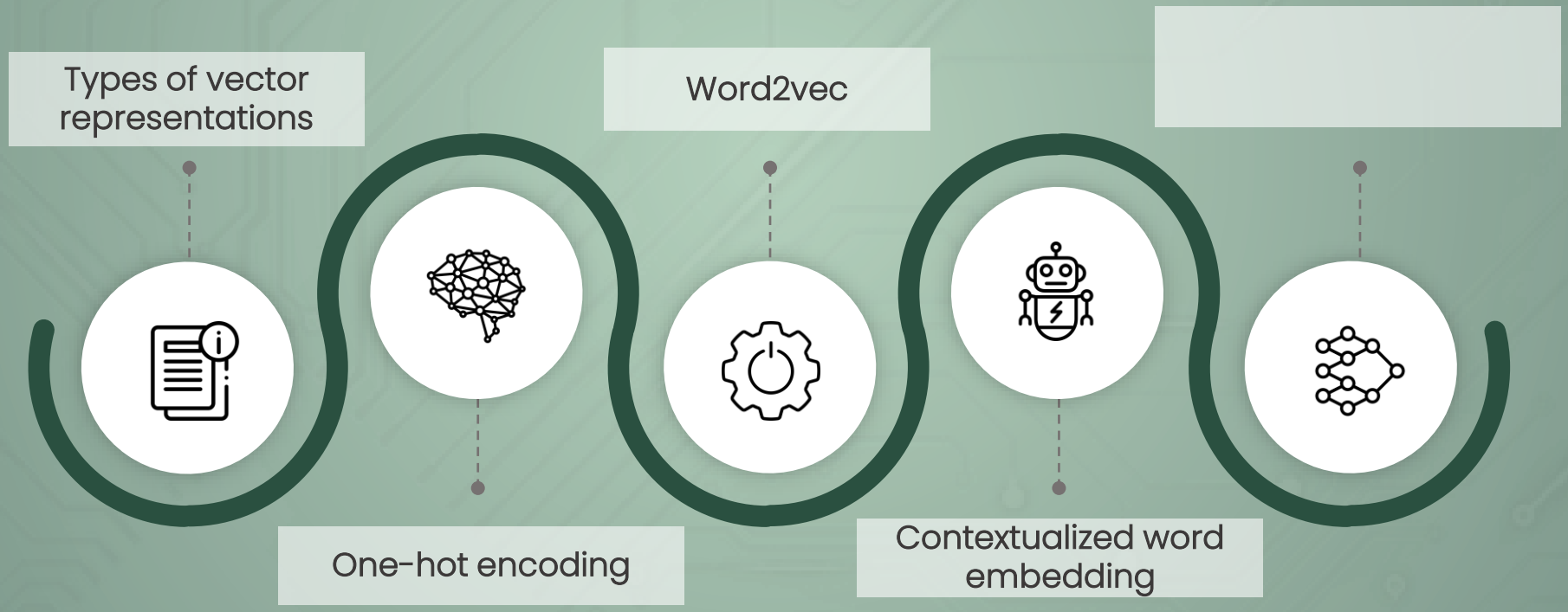
Machine Learning:
Applies algorithms for interpretation. Ranges from rule-based to deep learning.

Output:
Generates responses or actions. Examples:

- Text generation
- sentiment classification.

Vector Representations

...involves the organization, summarization, and visualization of data. It provides simple summaries about the sample and the measures.



Vector representations

...enable us to convert textual data into numerical forms that can be processed by machine learning models.

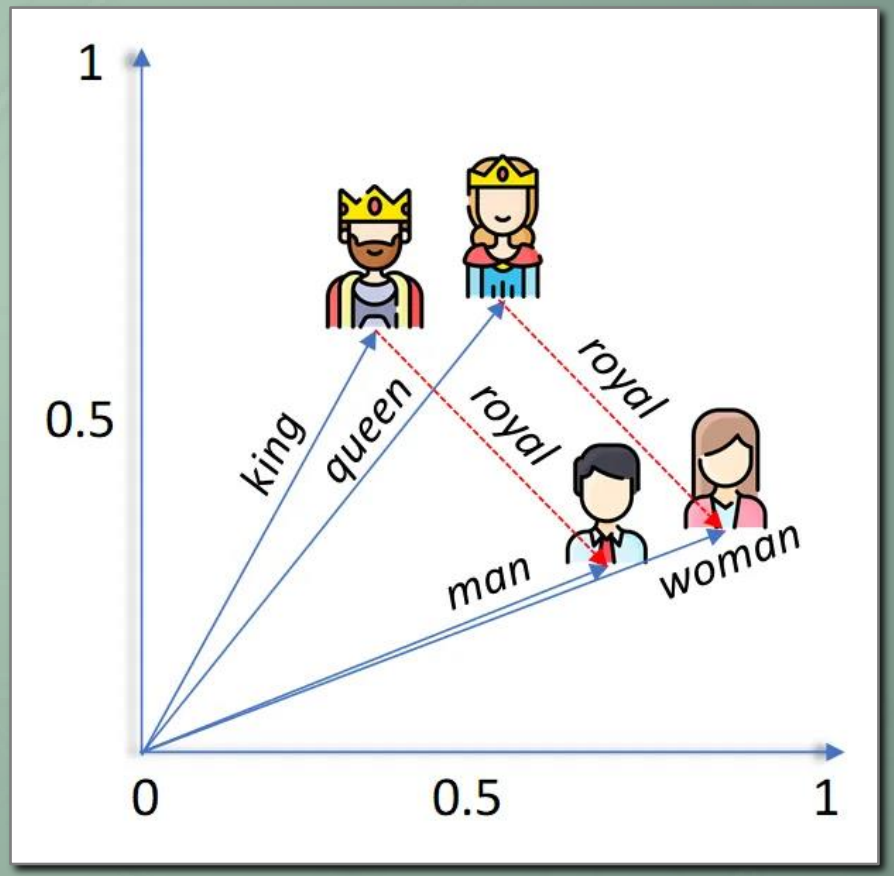
Three common methods of vector representation include

One-Hot Encoding:
Learns by interacting with an environment.

Word2Vec:
Provides feedback to the agent.

Contextualized Word Embedding:
Represents the policy or value function, guiding the agent's decisions.

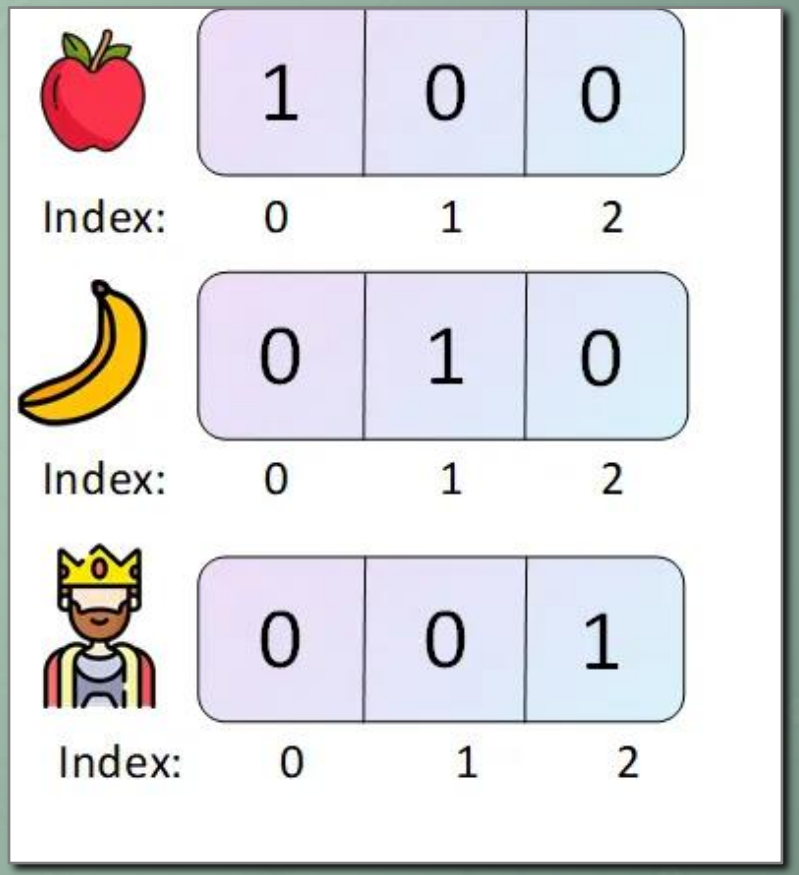
These methods play a crucial role in capturing semantic meaning and relationships between words and phrases.



One-hot encoding

...are fundamental in Natural Language Processing (NLP)

- basic method for vectorizing words in NLP.
- Each word in a vocabulary is represented as a binary vector:
  - A vector of all zeros except for a single 1
  - indicating the word's presence.
- Simple and intuitive, but has limitations:
  - Doesn't capture semantic relationships between words.
  - High dimensionality in large vocabularies.
- Often used as a starting point for more advanced techniques



Word2Vec

...is a popular word embedding technique that represents words in a continuous vector space.

Key Features:

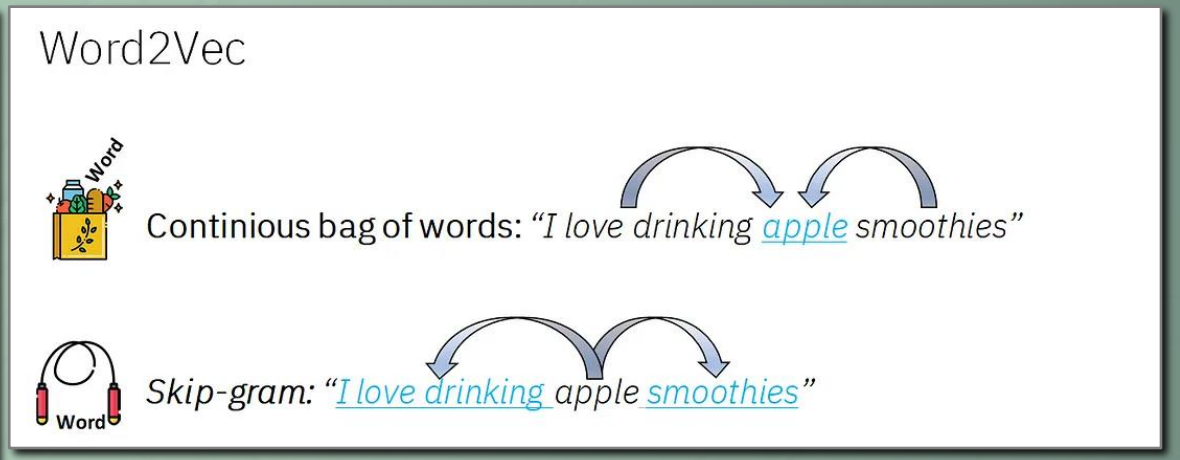
- Captures semantic meaning
- Words with similar meanings are closer in the vector space

Two Training Methods:

- Continuous Bag of Words (CBOW): Predicts a word given its context
- Skip-Gram: Predicts context words from a given target word

Benefits:

- Enables better performance in NLP tasks
- Helps in capturing semantic relationships (e.g., "king" - "man" + "woman" ≈ "queen")



Contextualized Word Embedding

Unlike traditional embeddings, contextualized embeddings generate word representations based on their specific context within a sentence, allowing for dynamic meanings.

Key Features:

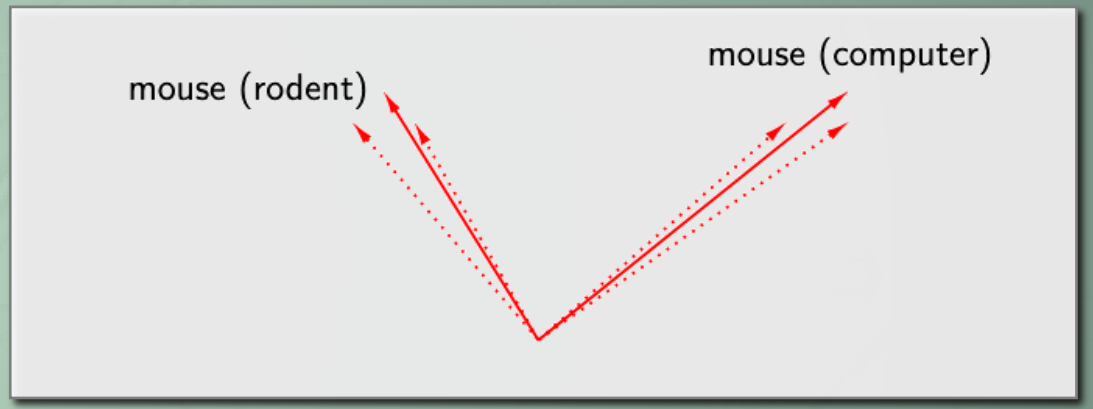
- Words can have different vectors in different contexts.
- Captures polysemy: a word's ability to have multiple meanings.

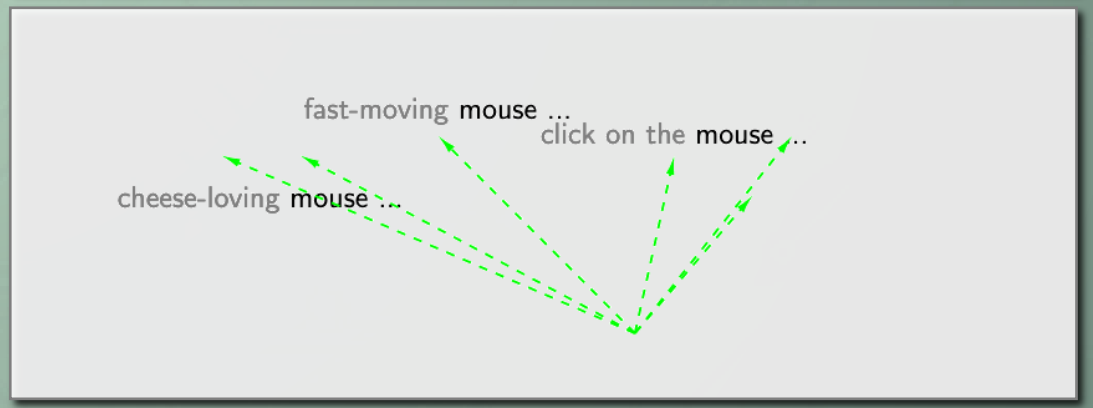
Popular Models:

- ELMo (Embeddings from Language Models)
- BERT (Bidirectional Encoder Representations from Transformers)

Benefits:

- Enhanced understanding of word nuances and meanings.
- Improved performance on downstream NLP tasks.

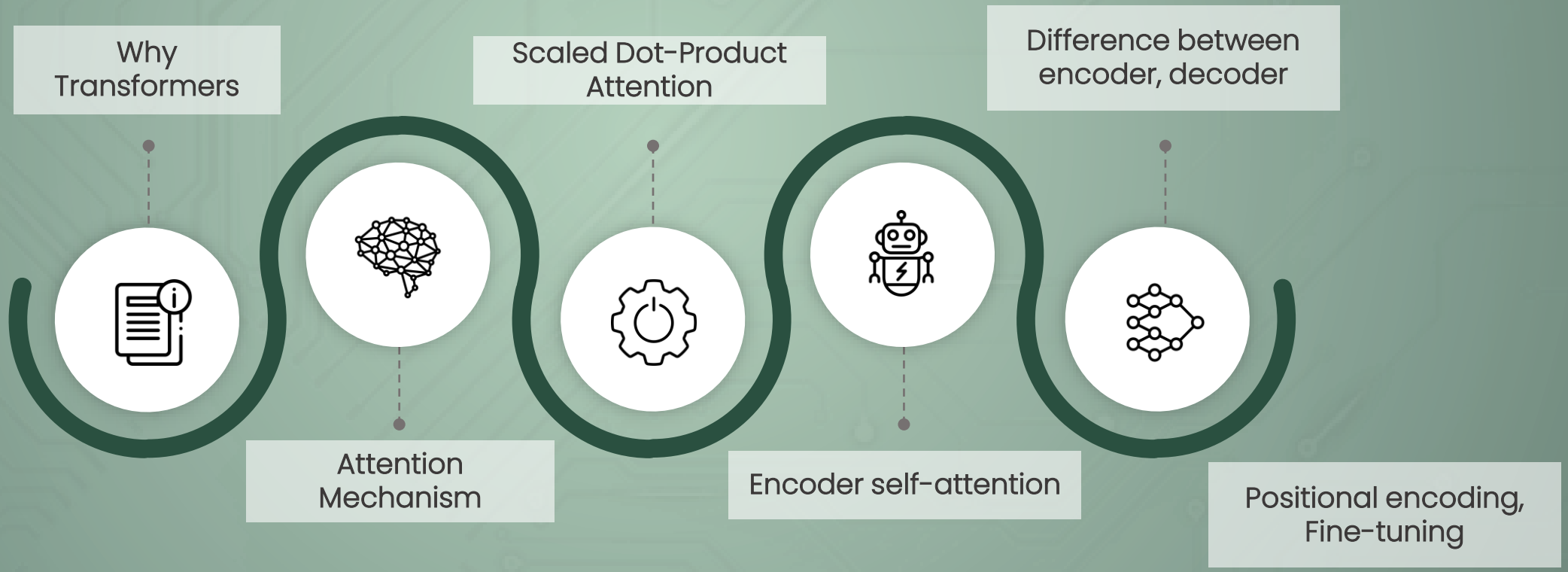




Transformers

...involves the organization, summarization, and visualization of data. It provides simple summaries about the sample and the measures.

Bridge to words embeddings, vector representations!!!



Why Transformers are significant

Transformers excel at modeling sequential data like natural language.
Nowadays encoders, decoders are used separately.

Encoder:
- Sentiment Analysis
- Text classification

Decoder:
- Conversation (ChatGPT)
- Translation

Comparison with RNNs:

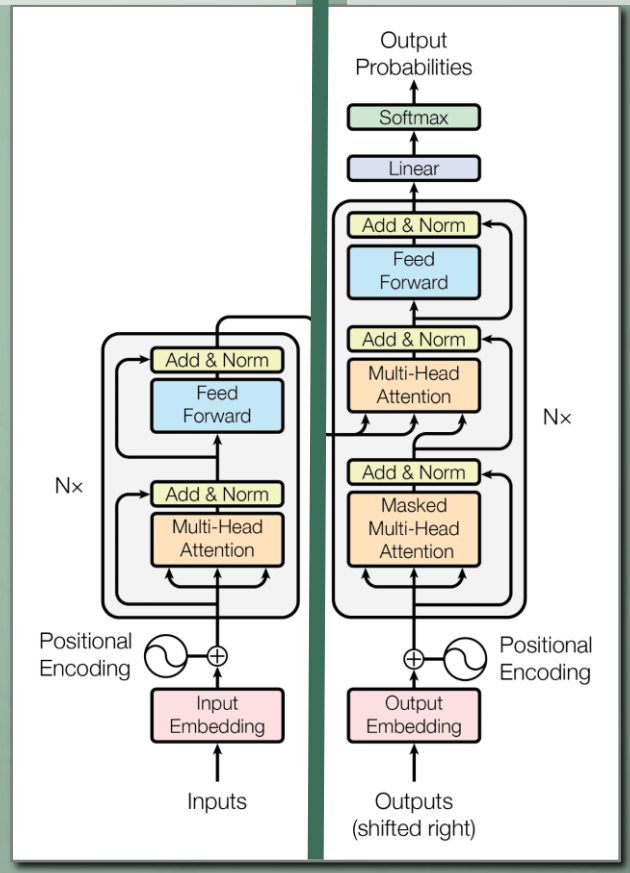
- Parallelizable and efficient on GPUs & TPUs.
- Replaces recurrence with attention for simultaneous computations.
- Outputs computed in parallel, unlike RNNs' series.

Advantages Over RNNs and CNNs:

- Captures distant or long-range contexts in data.
- Connects distant positions in sequences for longer connections.
- Uses attention to access entire input at each layer, unlike RNNs, CNNs.

Unique Characteristics:

- No assumptions on temporal/spatial relationships.
- Ideal for processing sets of objects (e.g., StarCraft units).



Attention Mechanism in NLP

A technique in deep learning models, especially in sequence-to-sequence tasks, that allows the model to focus on specific parts of the input when producing an output.

Key Features:

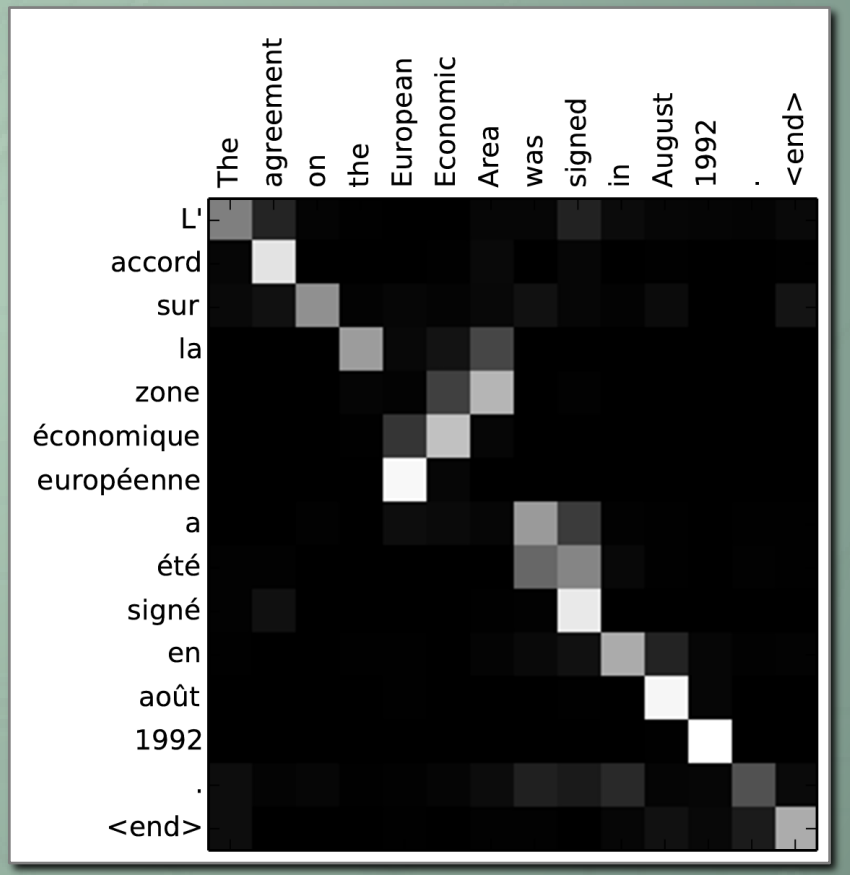
- Dynamically weighs input elements.
- Enhances the capturing of long-range dependencies in sequences.

Usage:

- Machine Translation: Helps in aligning words in source and target languages.
- Text Summarization: Prioritizes crucial parts of the content.

Benefits:

- Improves model's ability to remember long sequences.
- Enhances accuracy in tasks like translation and summarization.



Scaled Dot-Product Attention

...is a mechanism used in attention models that calculates attention scores based on dot product of query and key, scaled down by square root of their dimensionality.

- Queries: derived from input data, represent focus of model
- Keys: also derived from input data, interpretation: "labels" for the input data
- Values: are weighted based on compatibility of query and corresponding key

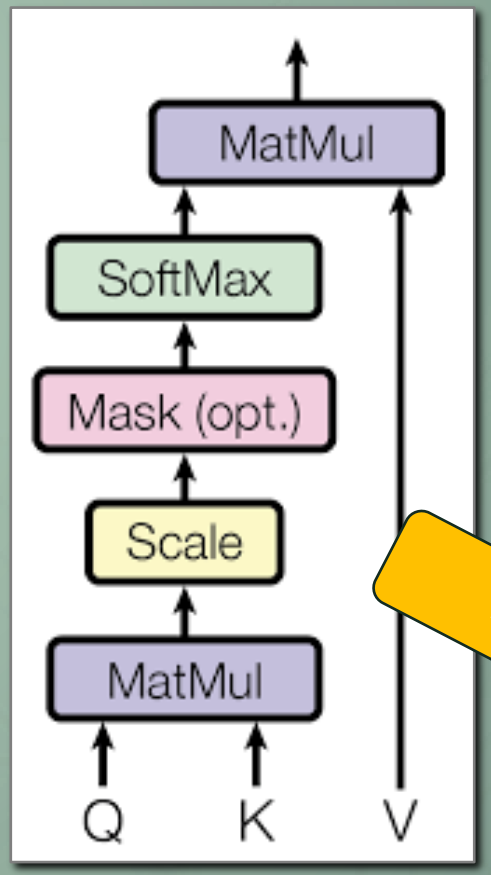
Formula

$$\boldsymbol{Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \times V}$$

- Q, K, V: Query, Key, and Value vectors respectively
- $\boldsymbol{d_k}$: dimensionality of query/key vectors

How it works

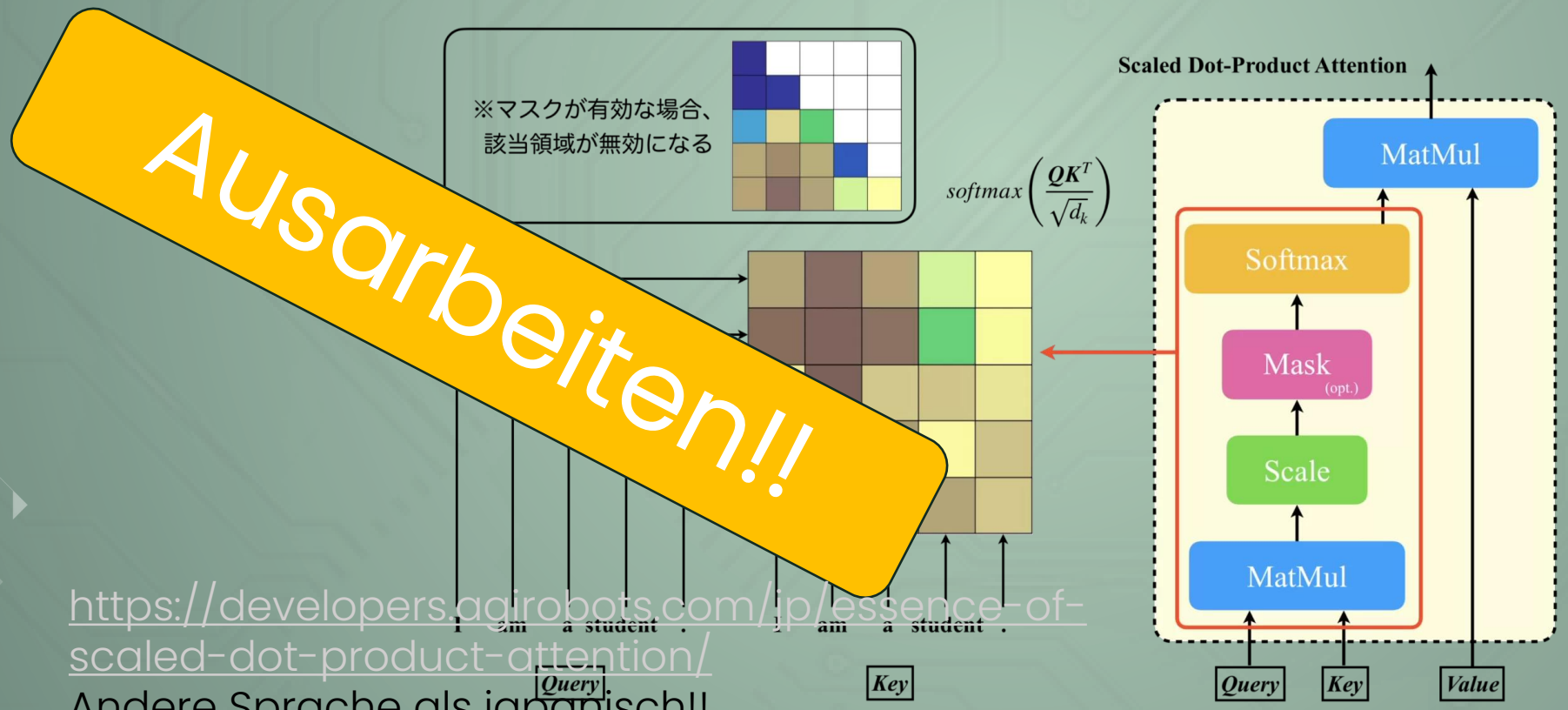
- For each query, attention mechanism computes score with each key in input. Score represents how well query aligns with particular key.
- Scores are used to create a weighted combination of the values.
- If key aligns well with query, value associated with that key gets larger weight in final output.



Ausarbeiten!!

Scaled Dot-Product Attention

...is a mechanism used in attention models that calculates attention scores based on dot product of query and key, scaled down by square root of their dimensionality.

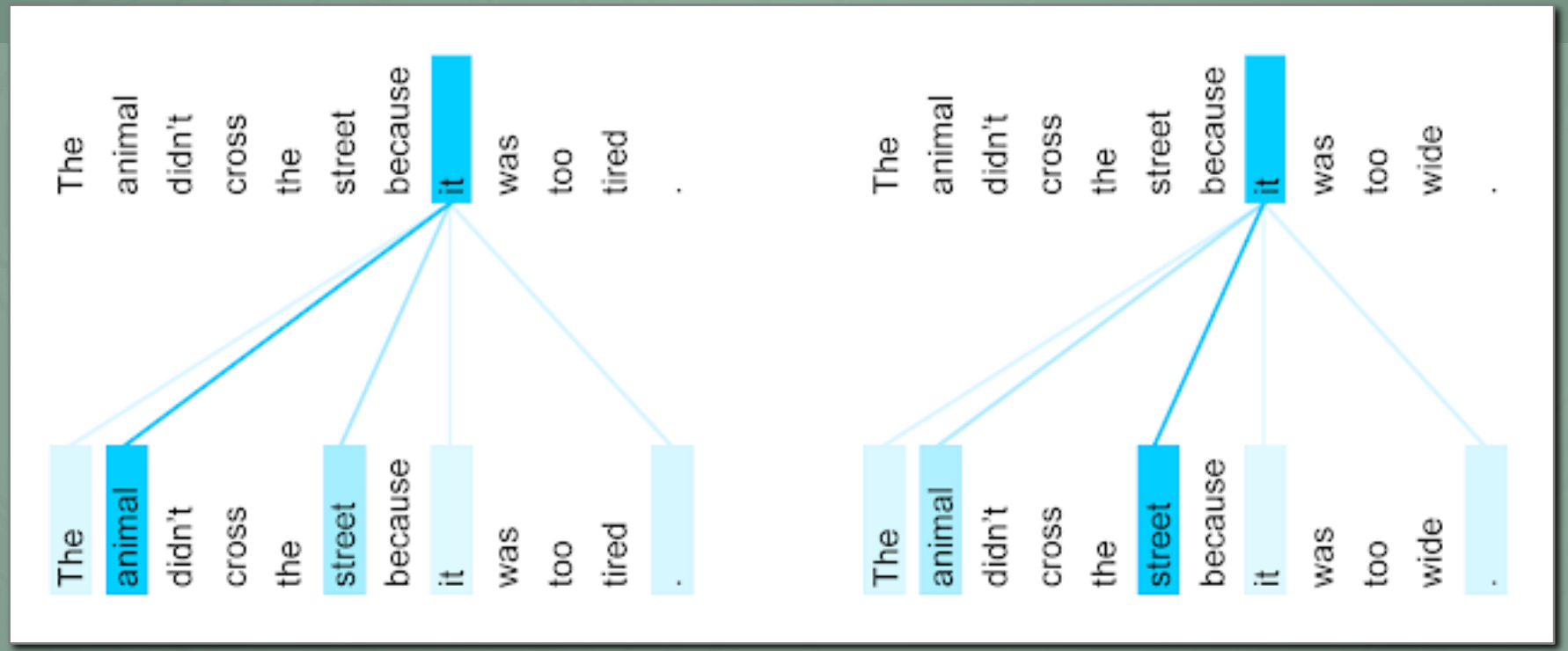


https://developers.agirobots.com/jp/essence-of-scaled-dot-product-attention/

Andere Sprache als japanisch!!

Encoder self-attention distribution: word “it”

Meaning of the word „it“ depends on context. Any complex task like translation or even Sentiment Analysis of multi-sentence comments needs context awareness.



“it” means animal

“it” means street

Difference between encoder, decoder

Initially, i.e. in 2014, transformer based translators have used both encoders and decoders. Encoder connectes words (tokens) in both directions, decoder predicts next word.

Positional Encoding

...are fundamental in Natural Language Processing (NLP)

Definition

itional Encoding is a crucial element in the Transformer
cture.

- Trans ledge of the order of words in a sequence
- Positional encod ation to word embeddings.
- How it works:

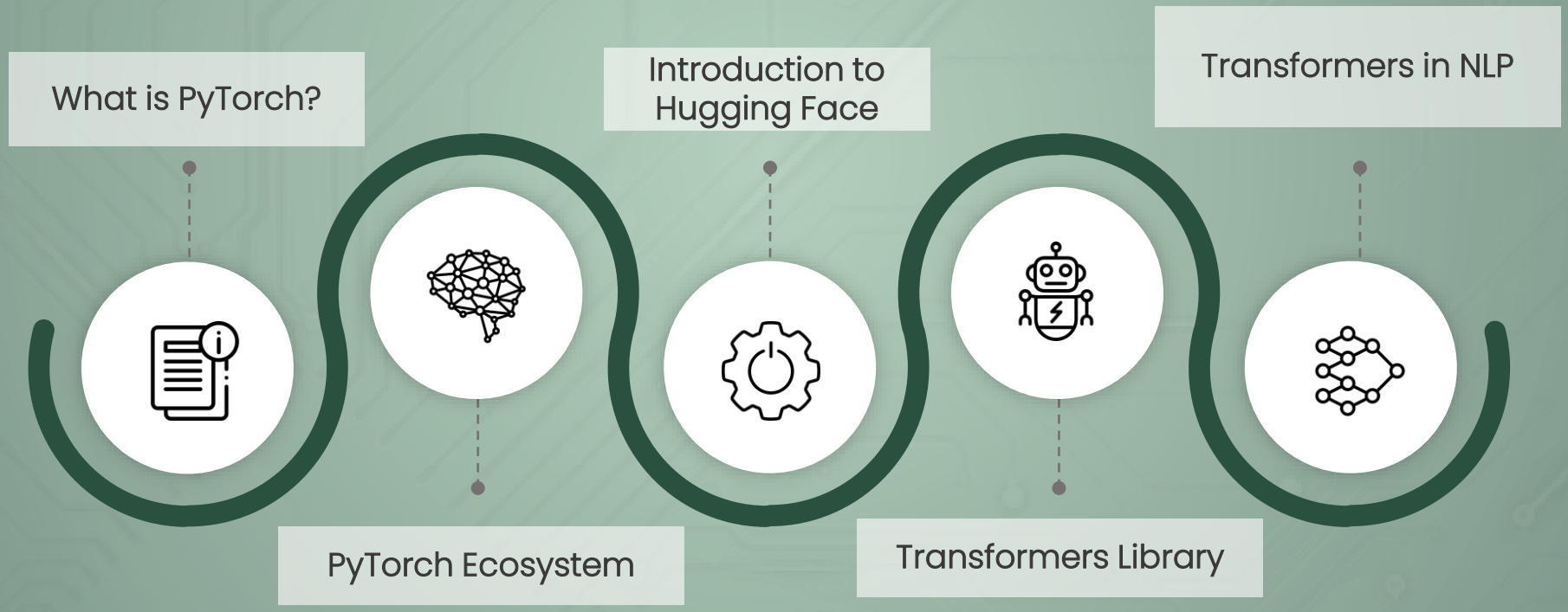
- Embeds the position of each word in t vector space.
- Helps the model understand the sequential aspe nput data.
- Positional encoding is essential for Transformers to handle sequences effectively, especially in tasks like language understanding and generation.

Ausarbeiten!!

Positional encoding,
Fine-tuning

Programming Tools

...involves the organization, summarization, and visualization of data. It provides simple summaries about the sample and the measures.



What is PyTorch?

It is an open-source deep learning framework developed by Facebook's AI Research lab.

Key Features:

- Dynamic computational graph, offering flexibility in model design.
- Native support for GPU acceleration, enhancing performance.
- Intuitive tensor library similar to NumPy but with GPU support.

Applications:

- Research prototyping, offering ease and speed.
- Building deep learning models, from neural networks to complex architectures.

Community & Ecosystem:

- Active community contributing to its growth.
- Rich ecosystem with pre-trained models, tools, and libraries.

PyTorch Ecosystem

...are fundamental in Natural Language Processing (NLP)

Definition

- In the Transformer architecture, the Decoder plays a pivotal role in generating output sequences.

Key responsibilities of the Decoder:

- Utilizes the attention mechanism to focus on different parts of the encoded input sequence.
- Generates the output sequence token by token, using both context from the Encoder and its own self-attention mechanism.
- Significance:

- Enables the model to perform tasks like machine translation, text summarization, and text generation.
- Combines information from the Encoder with its own understanding of the output sequence structure.
- The Decoder is a fundamental component in many sequence-to-sequence tasks in NLP and has led to significant advances in language generation.

Introduction to Hugging Face

...are fundamental in Natural Language Processing (NLP)

~~Definition~~

- In the Transformer architecture, the Decoder plays a pivotal role in generating output sequences.

~~Key responsibilities of the Decoder~~:

- Utilizes the attention mechanism to focus on different parts of the encoded input sequence.
- Generates the output sequence token by token, using both context from the Encoder and its own self-attention mechanism.
- Significance:

- Enables the model to perform tasks like machine translation, text summarization, and text generation.
- Combines information from the Encoder with its own understanding of the output sequence structure.
- The Decoder is a fundamental component in many sequence-to-sequence tasks in NLP and has led to significant advances in language generation.

Transformers Library

...are fundamental in Natural Language Processing (NLP)

Definition

- In the Transformer architecture, the Decoder plays a pivotal role in generating output sequences.

Key responsibilities of the Decoder:

- Utilizes the attention mechanism to focus on different parts of the encoded input sequence.
- Generates the output sequence token by token, using both context from the Encoder and its own self-attention mechanism.
- Significance:

- Enables the model to perform tasks like machine translation, text summarization, and text generation.
- Combines information from the Encoder with its own understanding of the output sequence structure.
- The Decoder is a fundamental component in many sequence-to-sequence tasks in NLP and has led to significant advances in language generation.

Transformers in NLP

...are fundamental in Natural Language Processing (NLP)

Definition

- In the Transformer architecture, the Decoder plays a pivotal role in generating output sequences.

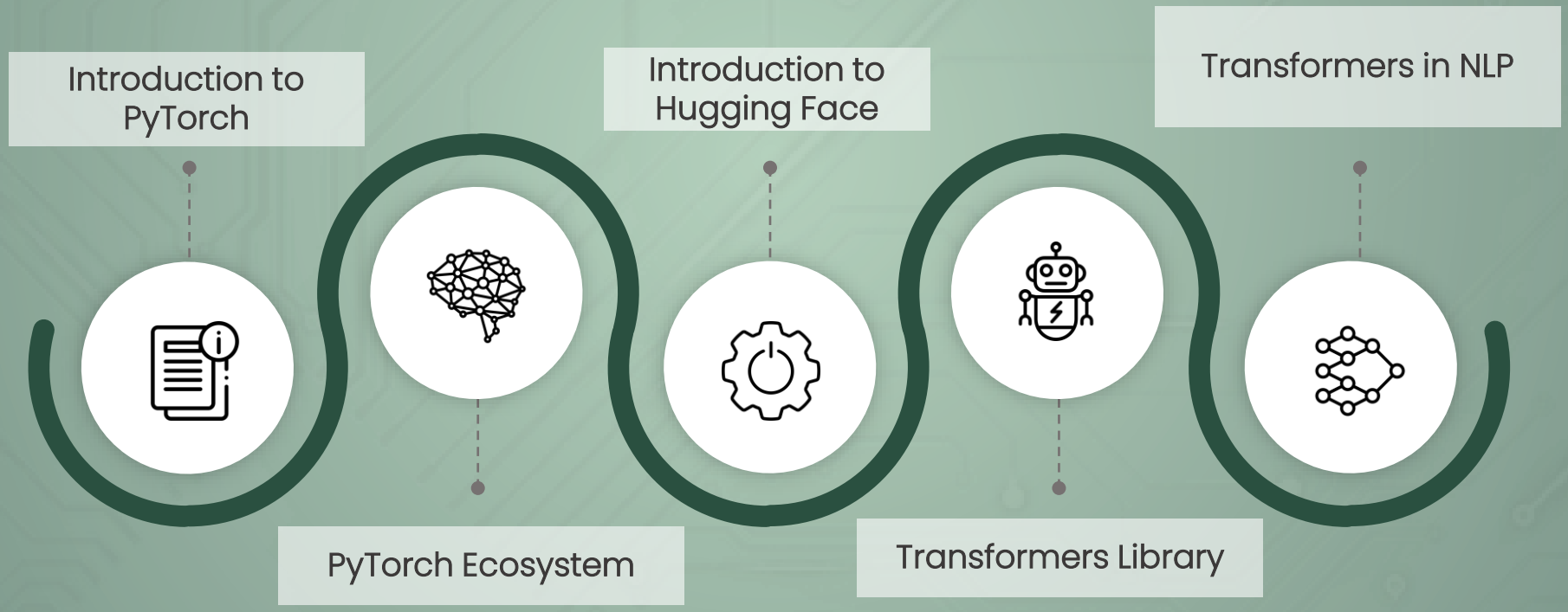
Key responsibilities of the Decoder:

- Utilizes the attention mechanism to focus on different parts of the encoded input sequence.
- Generates the output sequence token by token, using both context from the Encoder and its own self-attention mechanism.
- Significance:

- Enables the model to perform tasks like machine translation, text summarization, and text generation.
- Combines information from the Encoder with its own understanding of the output sequence structure.
- The Decoder is a fundamental component in many sequence-to-sequence tasks in NLP and has led to significant advances in language generation.

Example: Sentiments of movie comments

...involves the organization, summarization, and visualization of data. It provides simple summaries about the sample and the measures.



ChatGPT/Dall-E3 Prompts

Illustration of a digital tablet lying flat, displaying a chat application with emojis representing various moods. Around the tablet, there are holographic projections of diverse faces showing emotions ranging from joy to sorrow. The entire scene is bathed in a gradient transitioning from green to gray.

Photo of a split screen showcasing on one side, a close-up of a computer screen displaying text inputs with highlighted positive and negative words, and on the other side, diverse faces of people with varying moods from happy to sad. Overlaying this scene, a translucent gradient flows from vibrant green at the top left corner to muted gray at the bottom right.

About me

Dr. Harald Stein

- Data Scientist ~ 7 years experience
- Algotrader ~ 4 years experience
- Ph.D. in Economics, Game Theory

- LinkedIn: https://www.linkedin.com/in/harald-stein-phd-1648b51a
- ResearchGate: https://www.researchgate.net/profile/Harald-Stein