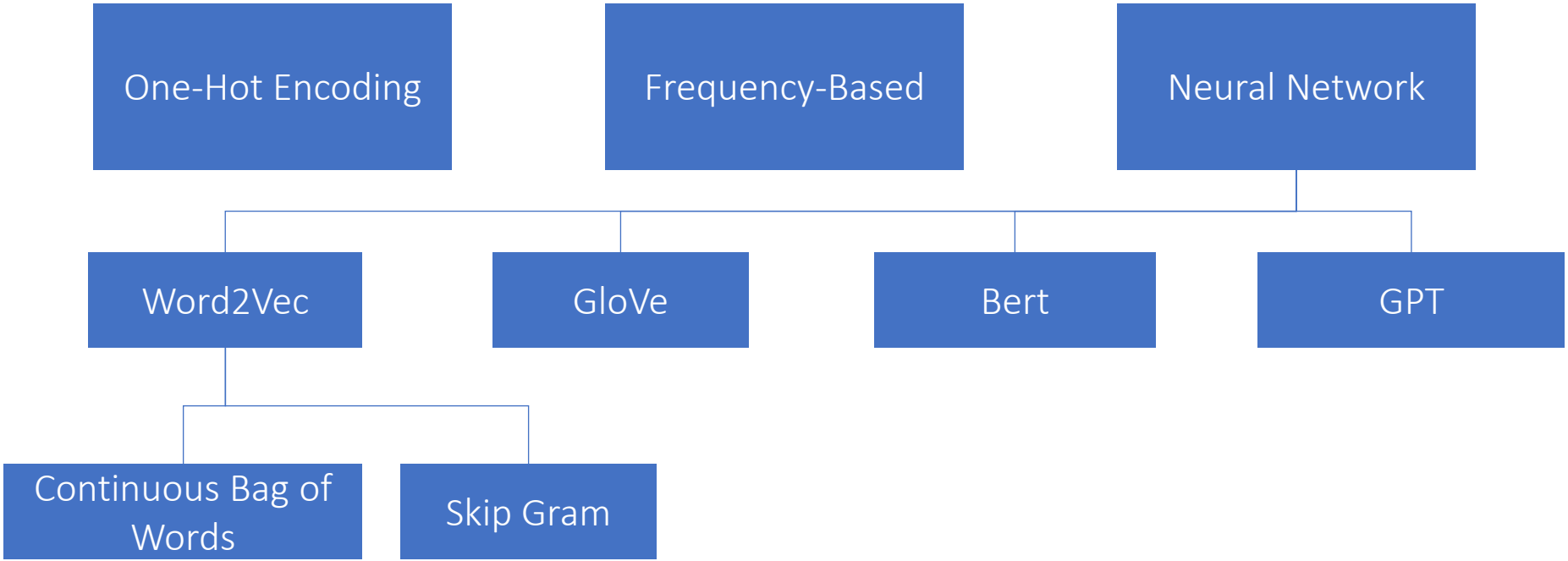
Word Embeddings based on Neural Networks

Word Embeddings represent words as **low-dimensional vectors** in mathematical space and **capture** their semantic and syntactic **meaning**.

Natural Language Processing

Word Embedding Approaches



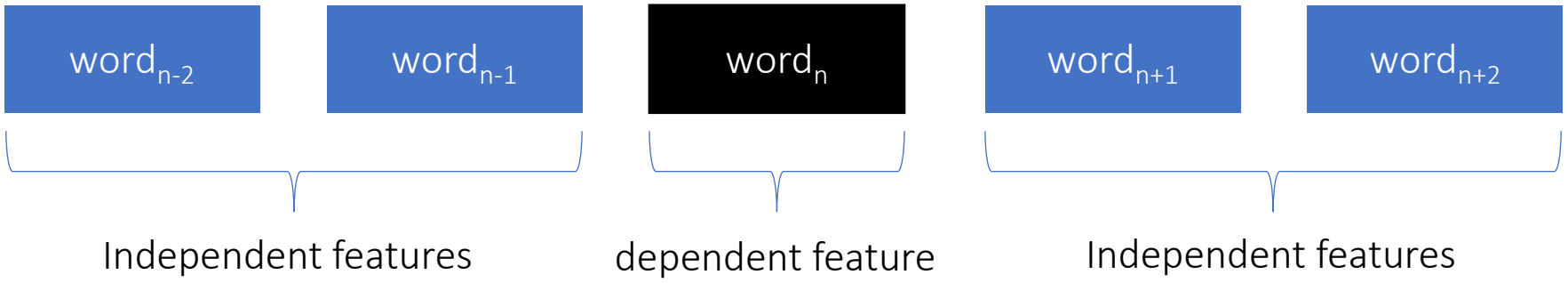
Natural Language Processing

Neural Network based Embeddings

- Aim to
  - Capture context / meaning
  - Capture similarity to other words
  - Reduce dimension
  - Avoid memory issues
- Developed based on Neural Networks

Natural Language Processing

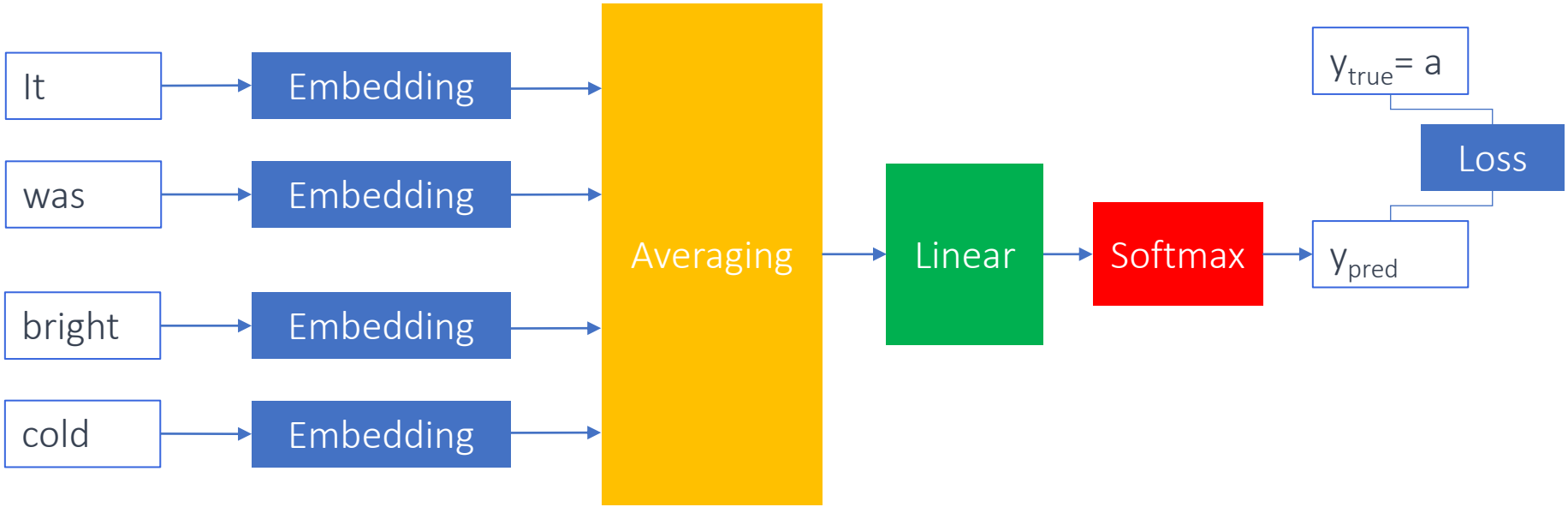
Word2Vec: Continuous Bag of Words



Independent Features	Dependent Feature“
[“It”, “was”, “bright”, “cold”]	“a”
[“was“, “a”, “cold”, “day”]	„bright“
…	

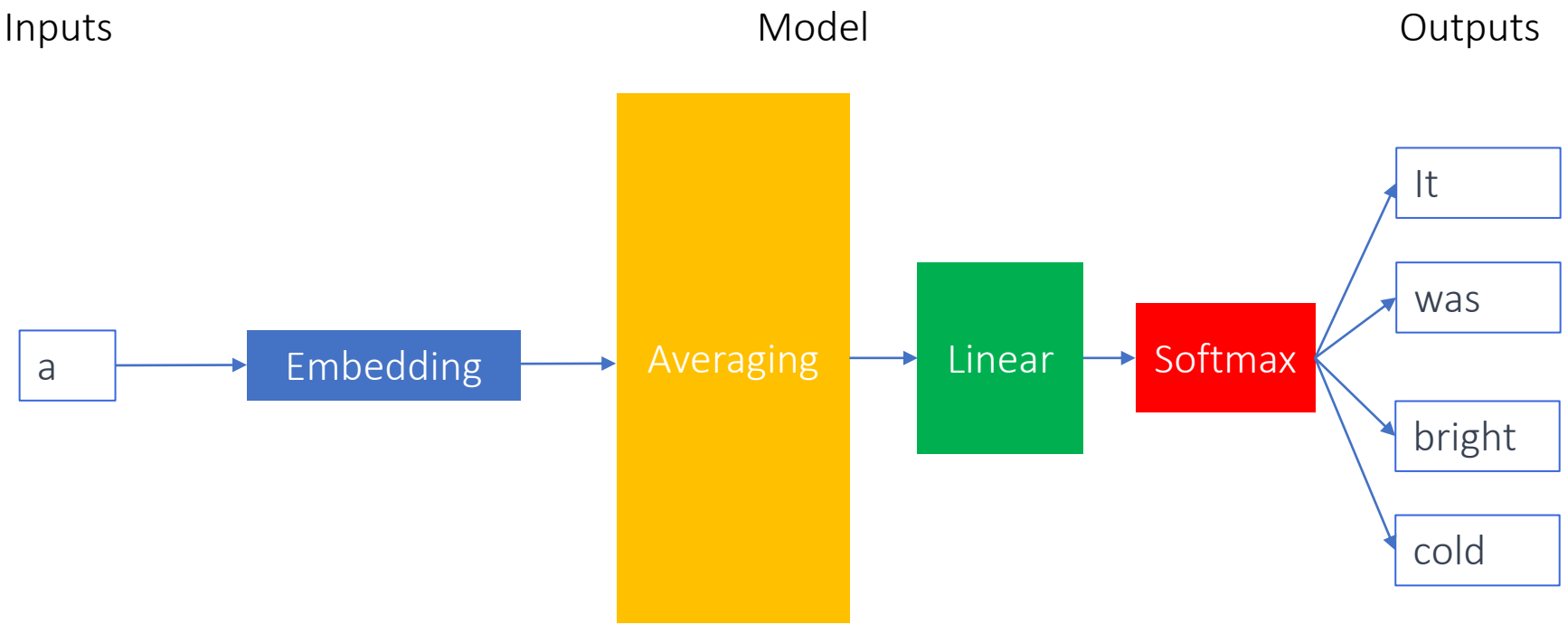
Natural Language Processing

Word2Vec: Continuous Bag of Words Model



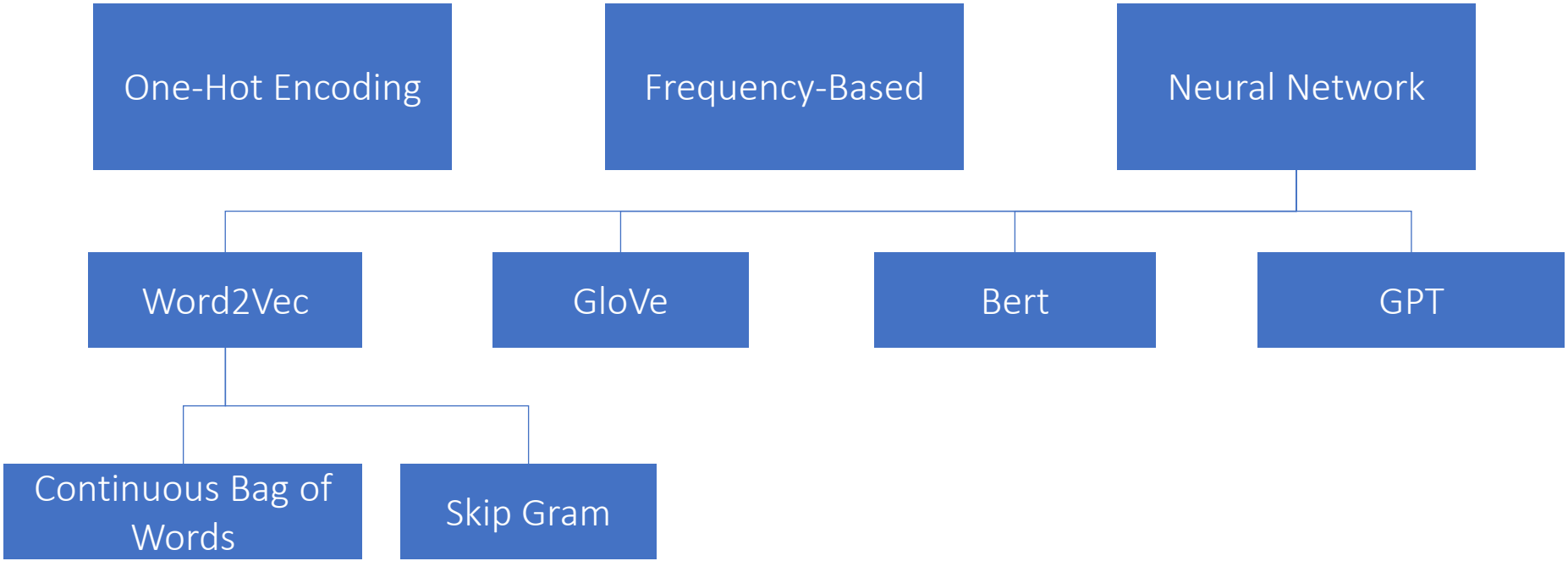
Natural Language Processing

Word2Vec: Skip Gram



Natural Language Processing

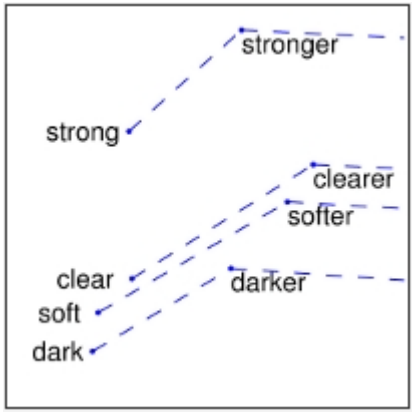
Word Embedding Approaches



Natural Language Processing

GloVe

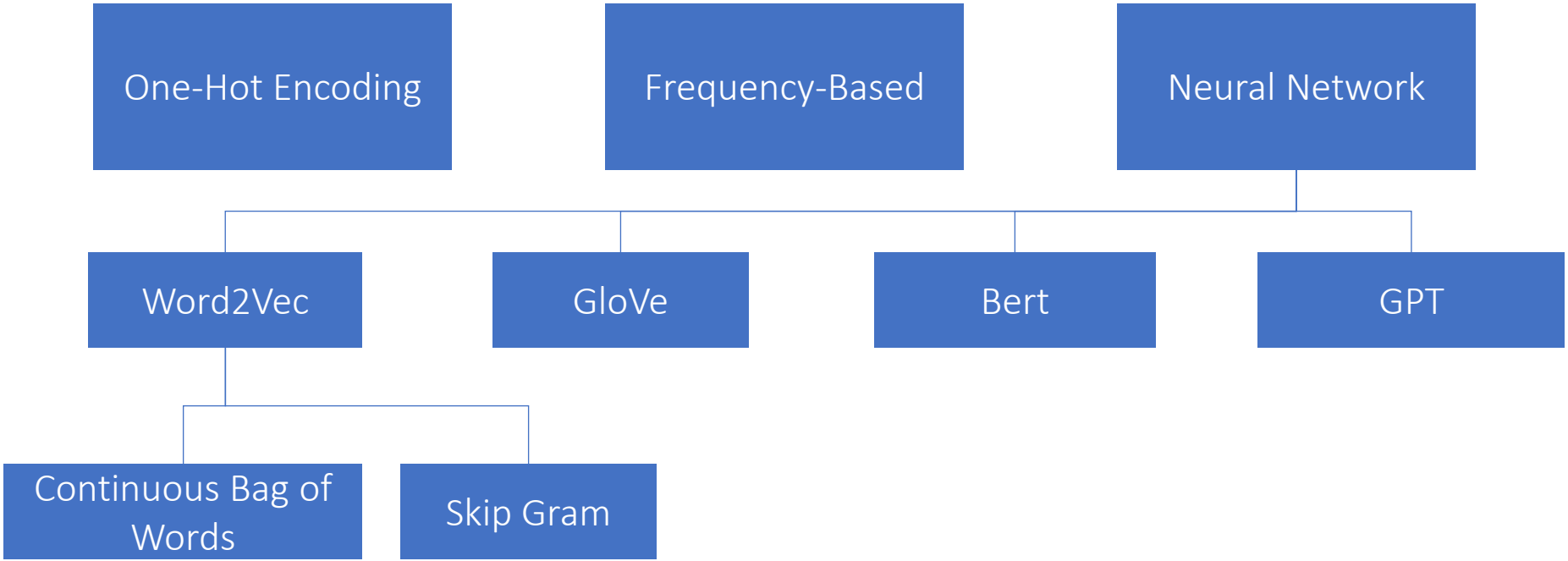
- Global Vectors for Word Representations
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation
- based on co-occurrence matrix of words in a corpus, which counts how often words appear together in the same context.
- constructs a matrix of word co-occurrence counts and then factorizes this matrix to obtain word embeddings
- factorization based on singular value decomposition (SVD)
- resulting embeddings are dense, low-dimensional vectors
- Encode words as vector of other words



Source: https://nlp.stanford.edu/projects/glove/

Natural Language Processing

Word Embedding Approaches



Natural Language Processing

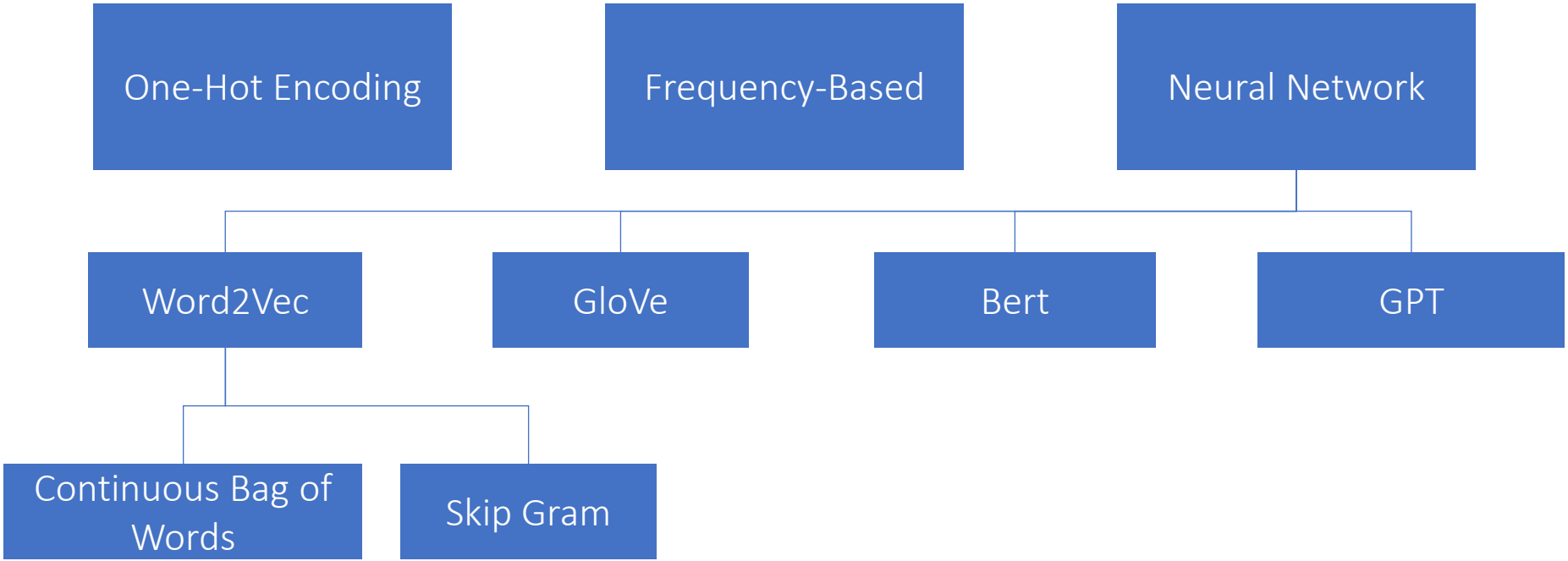
BERT

- **B**idirectional **E**ncoder **R**epresentations from **T**ransformers
- Developed by Google in 2018
- Pre-trained word embedding
- Based on Transformers
- Applies „masked language modeling“ – masking some words in sentence and learn to predict them
- Applies „next sentence prediction“ – model predicts whether two sentences are similar in a text

- Original variants: BERT-base (110m parameters, 440MB) and BERT-large (340m parameters, 1.3GB)
- Other variants: RoBERTa, ALBERT, ELECTRA, …

Natural Language Processing

Word Embedding Approaches



Natural Language Processing

GPT

- Generative Pre-trained Transformers
- Developed by OpenAI
- Not strictly a word embedding, but contextualized word embedding
- Unique embedding for each occurrence of a word based on surrounding words in text
- Applies Transformer architecture
- GPT-3 has 175 billion parameters