The birth of LLMs
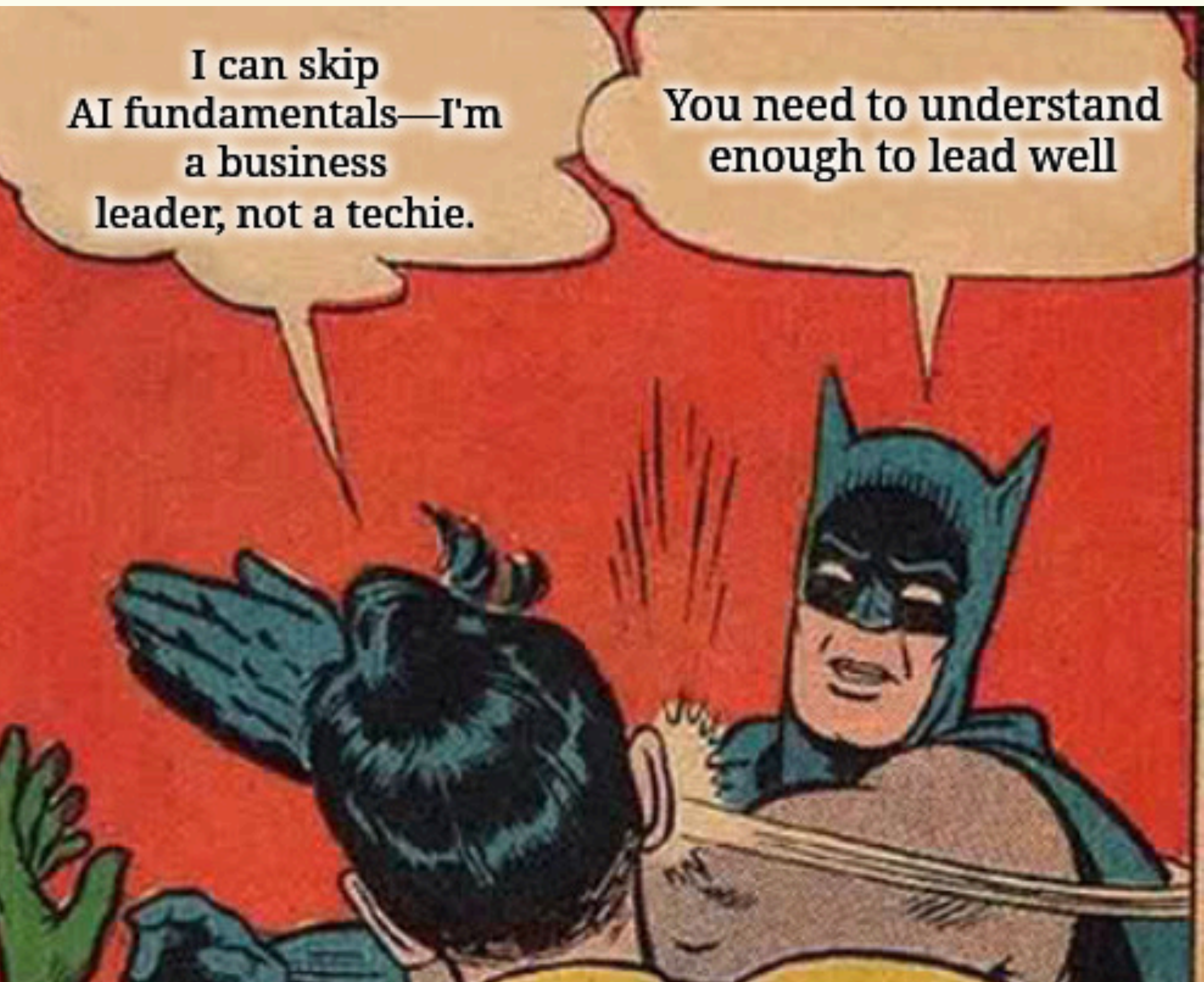
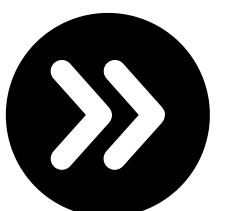📄 10 publications to master NLP history

👶 Where it All Began

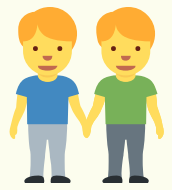📄 *A Mathematical Theory of Communication* **(Shannon, 1948)**

First mention of **n-grams**, a way to break text into sequences of words, making it easier for computers to understand language patterns

unigrams *This, is, a, sentence*
bigrams *This is, is a, a sentence*
trigrams *This is a, is a sentence*

👭 RNNs from The AI Godfathers

📄 *Neural networks and physical systems with emergent collective computational abilities.* **(Hopfield, 1982)**
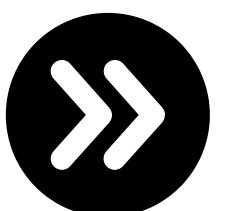
📄 *The Appeal of Parallel Distributed Processing* **(Hinton et al, 1986)**

a) Networks understanding sequences, making it possible for computers to understand context in text

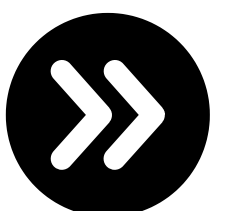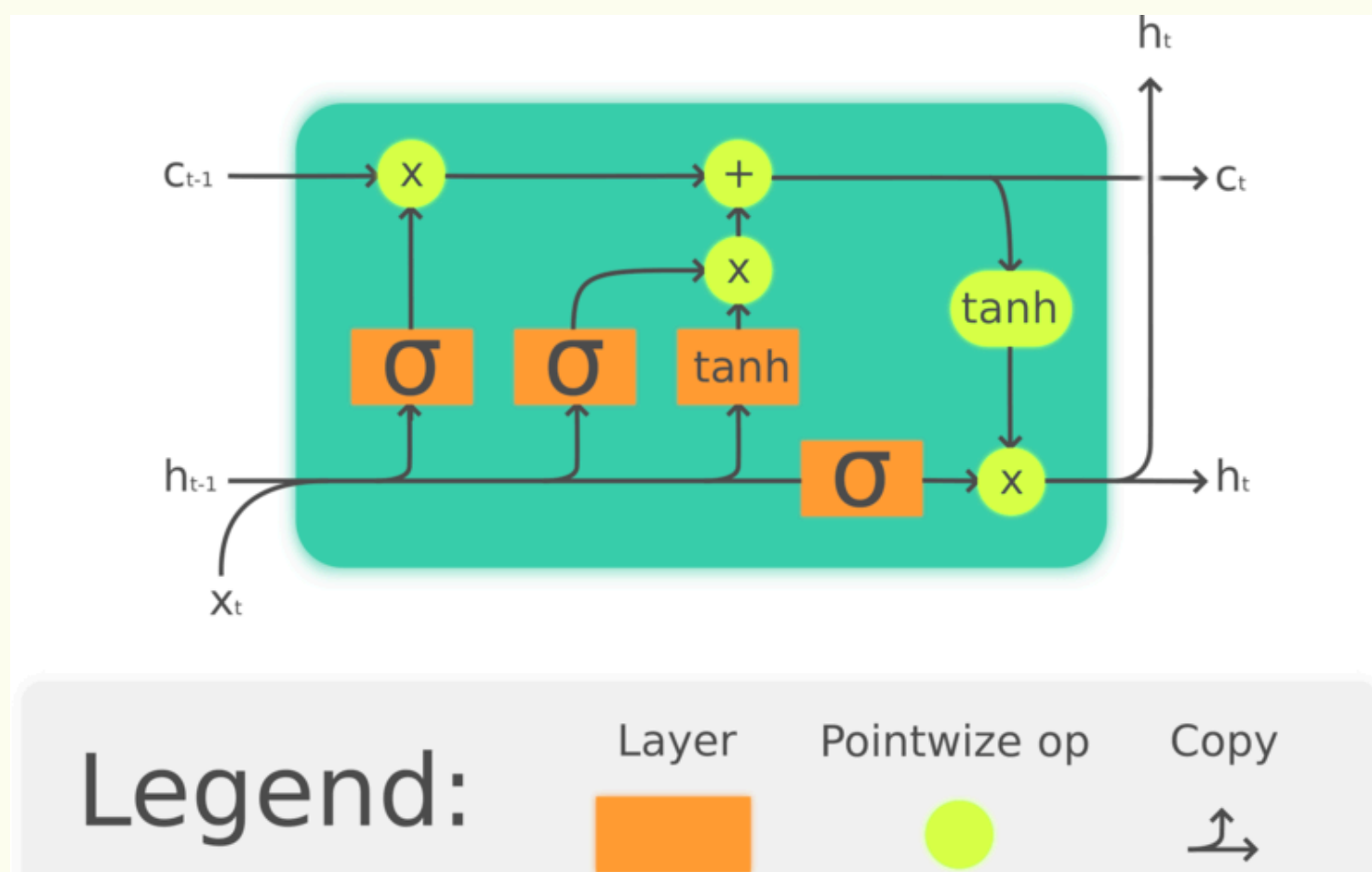b) First mentions of the idea of representing words as vectors

#panda-ai-pills

💾 LSTM

📄 *Long-Short Term Memory*
(Hochreiter et al., 1997)

Made it possible for models to remember important information over longer texts, helping computers understand meaning in longer pieces of writing.
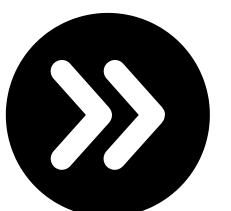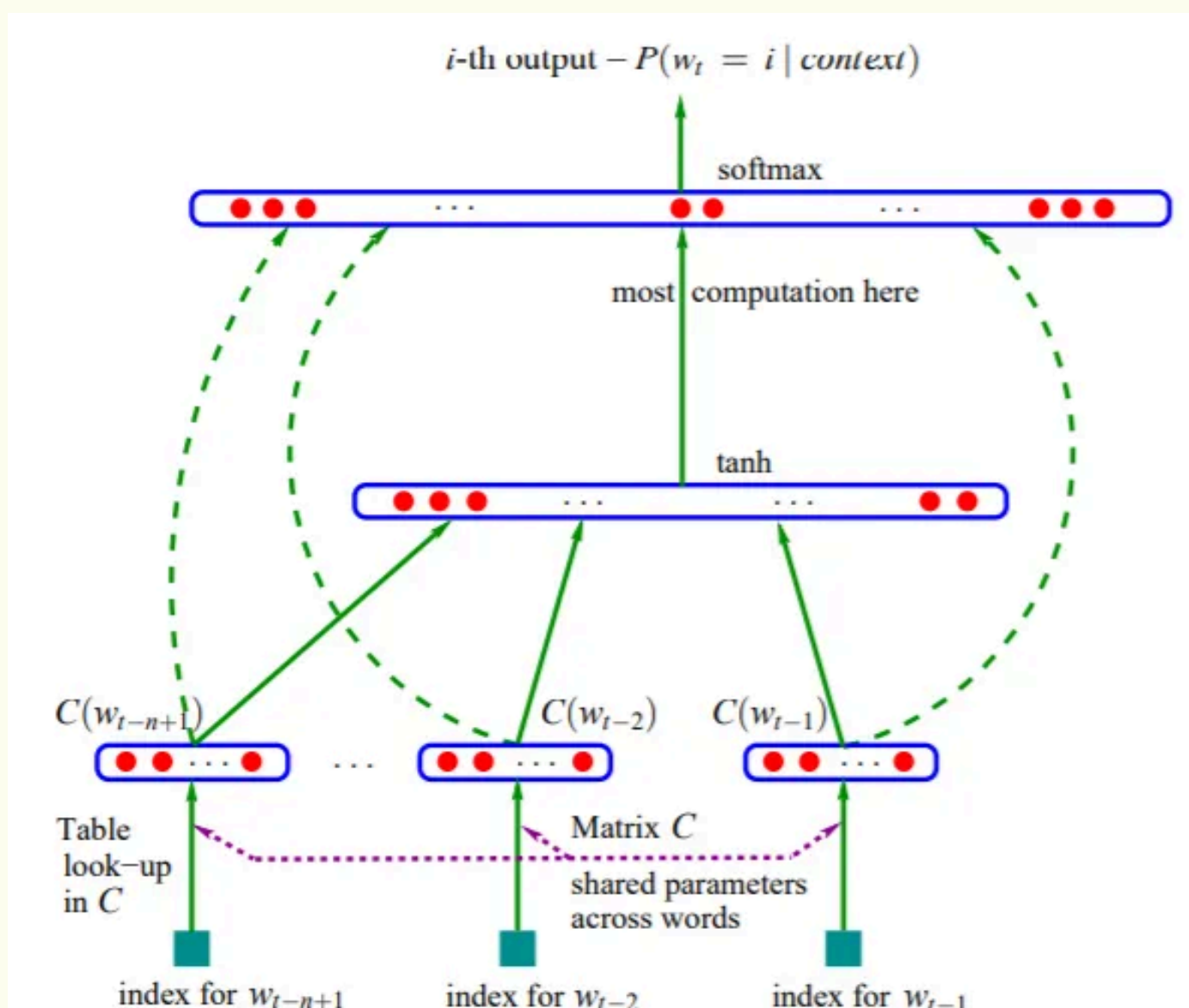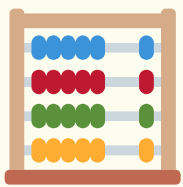
The 1st Language Model

📄 *A Neural Probabilistic Language Model* **(Bengio et al., 2003)**

Built the first model that used neural networks to predict the next word in a sequence. Words embedded as feature vectors
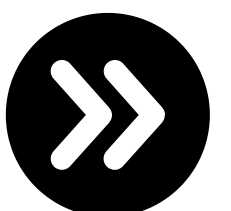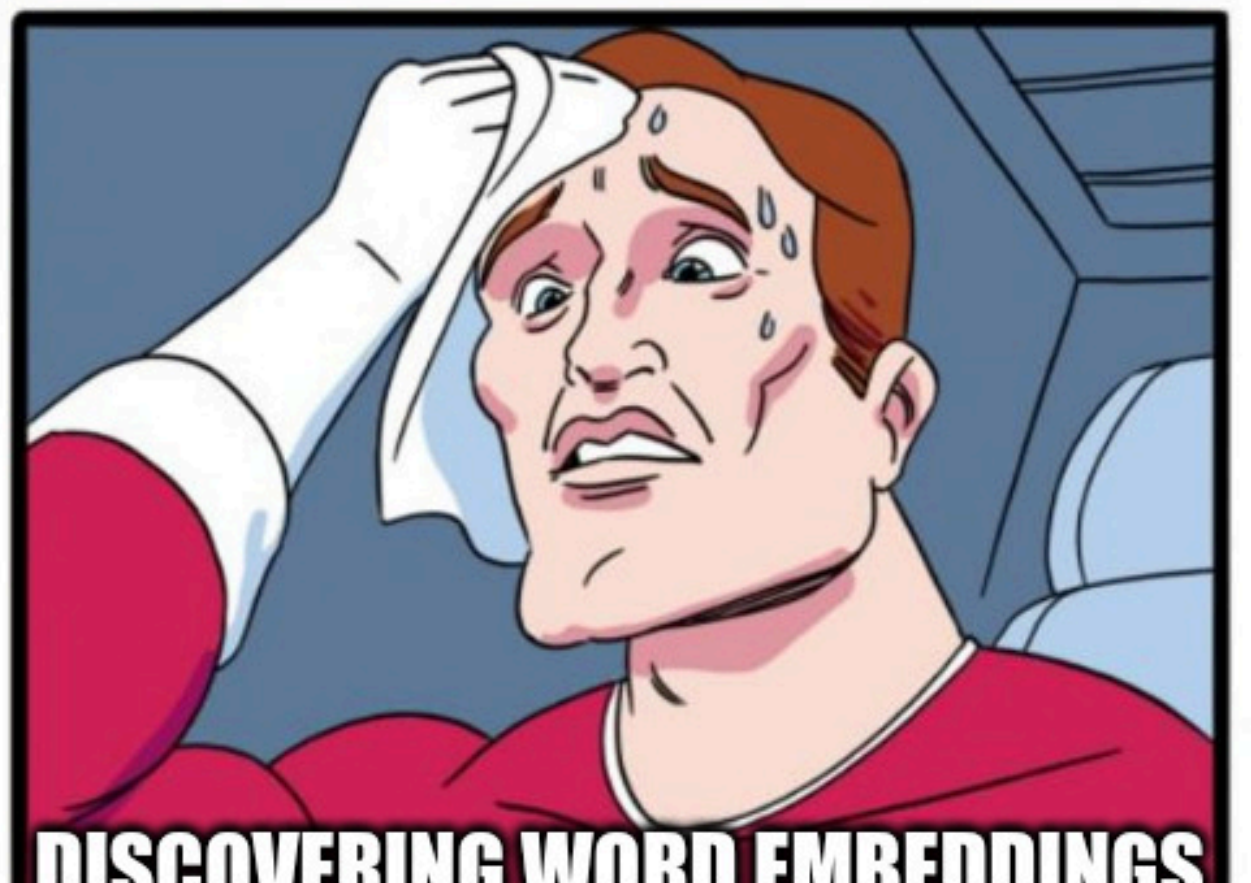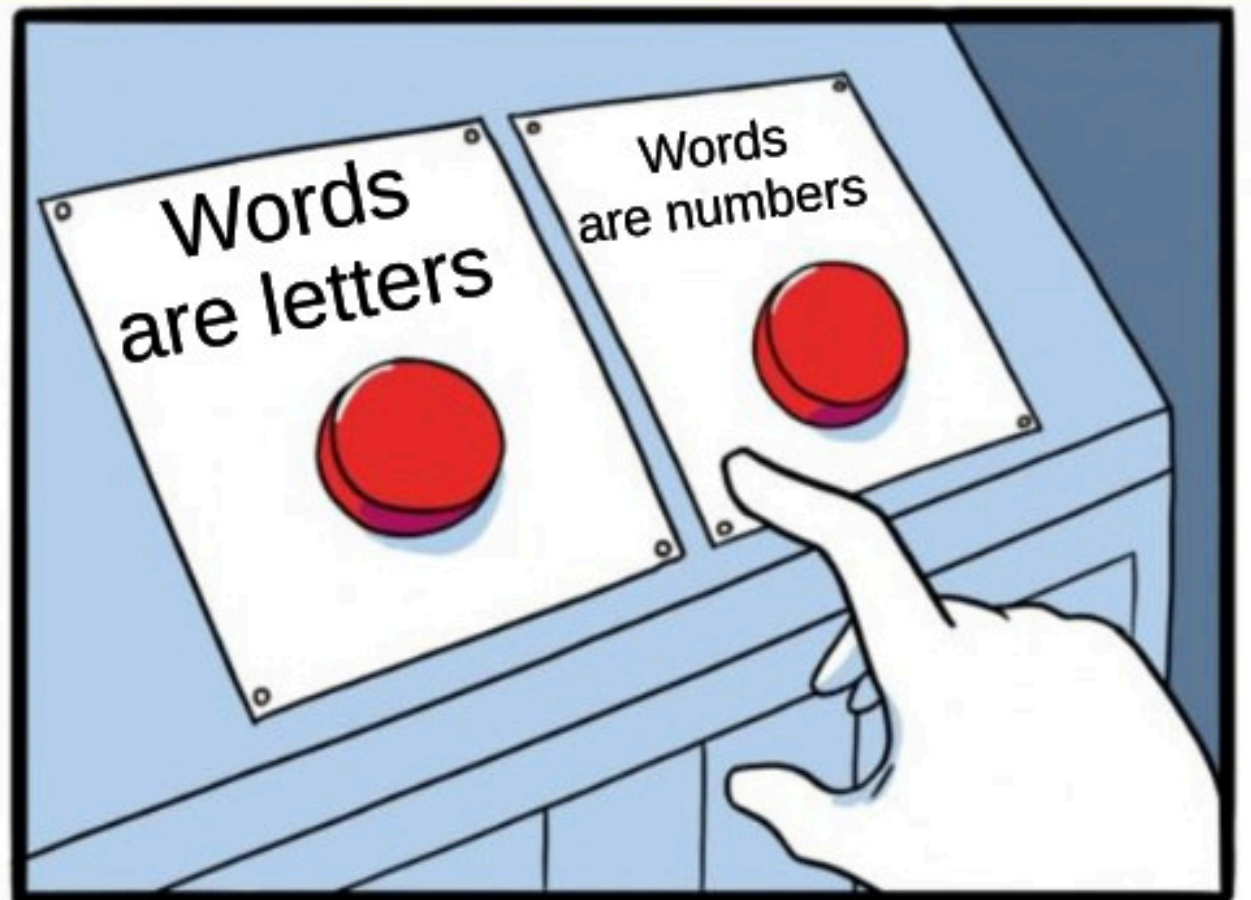
🧮 Words as Vectors

📄 *Efficient Estimation of Word Representations in Vector Space* (Mikolov et al., 2013)

Created a way to turn words into numbers that capture their meanings
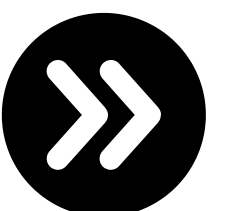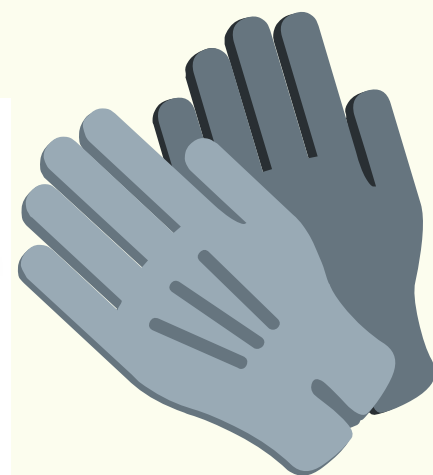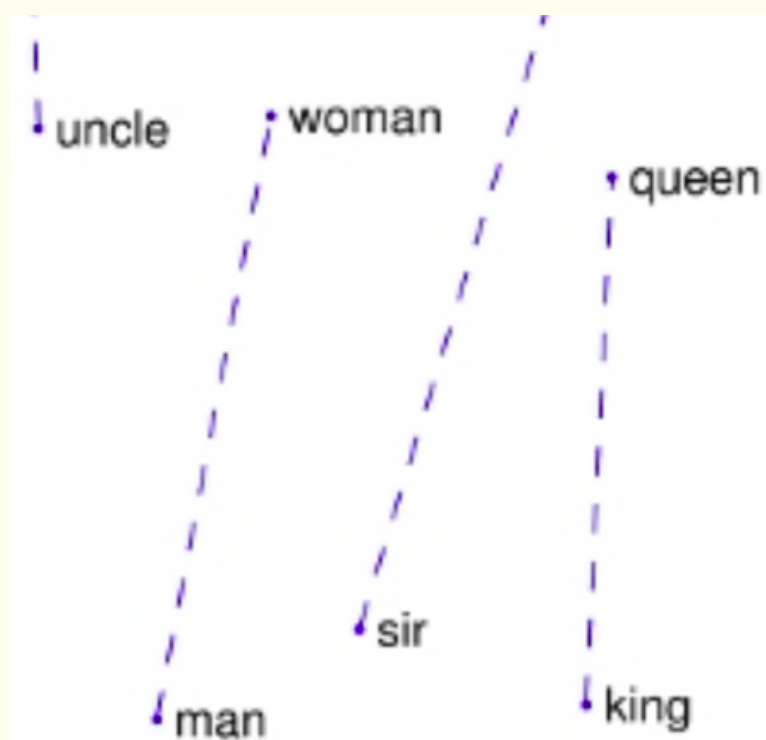
GloVe Embeddings

📄 *GloVe: Global Vectors for Word Representation* **(Pennington et al., 2014)**

Developed a method to understand word meanings by looking at how often words appear together (*co-occurence*), improving how computers understand relationships between word

🧐 Attention Mechanism

📄 *Neural Machine Translation by Jointly Learning to Align and Translate* (Bahdanau et al. 2015)
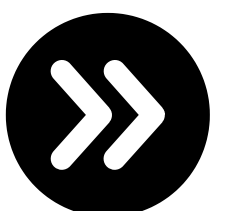
Created a mechanism to allow models to "remember" only the words that best helped it to translate a word correctly

Exercise improves physical health, boosts mood, and increases energy levels.

Attention
👇

Exercise improves physical **health**, boosts **mood**, and increases **energy** levels.
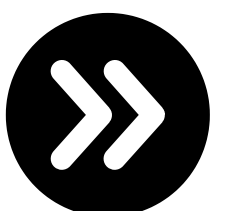
🤖 The Transformers

📄 *Attention is all you need (Vaswani et al. 2017)*

A new model architecture that uses only **attention**, without recurrence or convolutions. The real game changer in NLP & the underlying LLM technology.
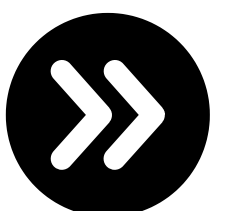
Embedding Language

📄 *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* **(Devlin et al. 2018)**

Do you remember Word2Vec and GloVe, which used the idea of mapping words into vectors to represent their meanings?

BERT took this concept further by using the Transformer architecture to map entire corpora, capturing the context of each word in both directions

Any question, doubt, or clarification?

Ask in comments