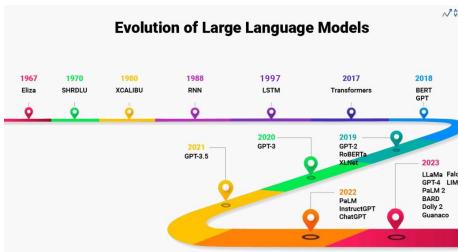


## Day 16: Language Models

### Introduction



What

Fundamental components of modern NLP

Basically, algorithms that understand and analyze human language

Increase in popularity after ChatGPT (LLM)

Question-answer

Summarization

Translation

Text-completion

Blog writing

Number of different tasks

A probability distribution over a sequence of words

In NLP it is used to predict the probability of a sequence of words in language

They are trained on "sets of sentences"

Then they learn the probabilities of different word combinations

Generate new text by predicting the probability of the next word in a sentence

To be precise, "an AI model which can understand and generate new text"

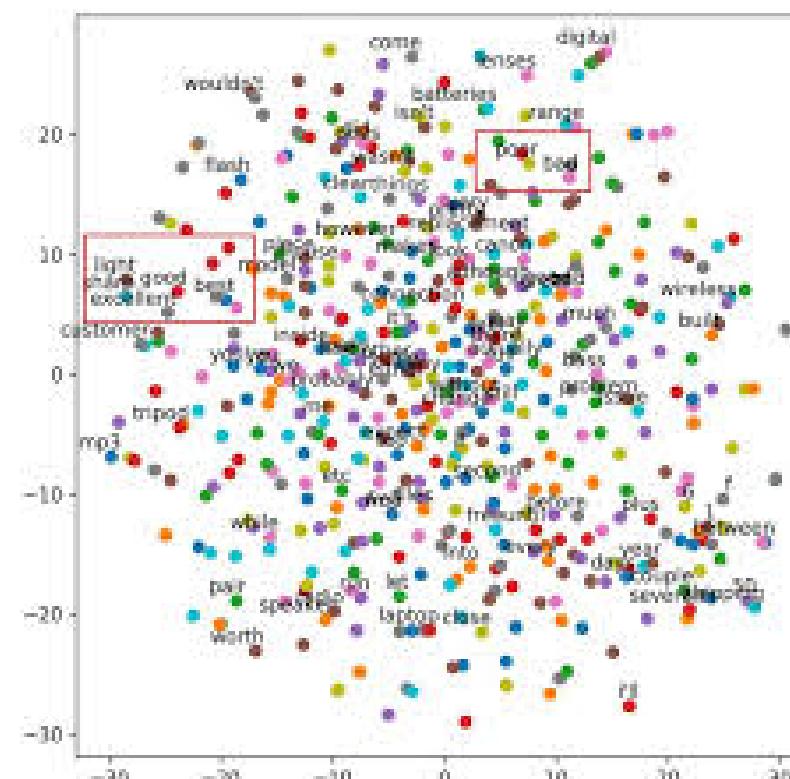
Main goal: Capturing statistical patterns

Eg: "I am" is followed by "eating"/"going" etc.

### Hands-on 1

#### Word Embeddings

A technique that represents words as numbers so that computer can understand them



It is done by representing words in the form of dense vectors of real numbers in a high dimensional space. (upto 1000)

Words which have similar meanings and used in similar contexts are assigned similar vectors i.e. located close to each other in the embedding space

Techniques - we use some pre-trained models to do this

Word2vec

GloVe

fastText

TF-IDF

### Hands-on 2