Introduction to NLP for Data Practitioners

Understanding Natural Language Processing (NLP)

Mihaela Danci
Data Analyst

linkedin.com/in/mihaela-danci/

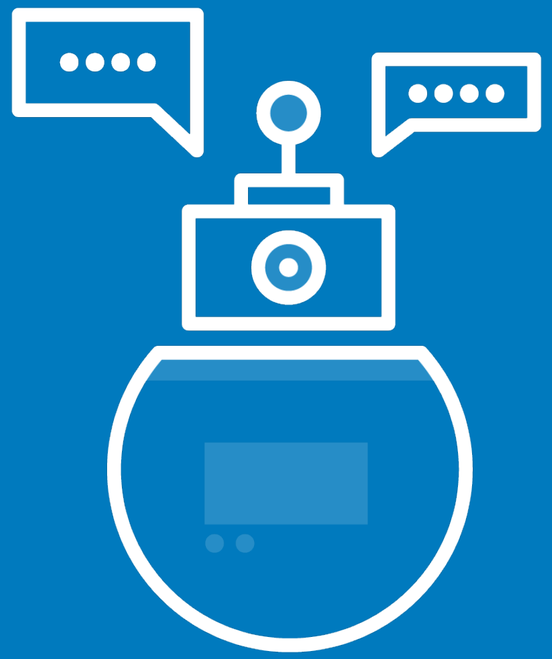
Language translation

Language translation

Spelling correction

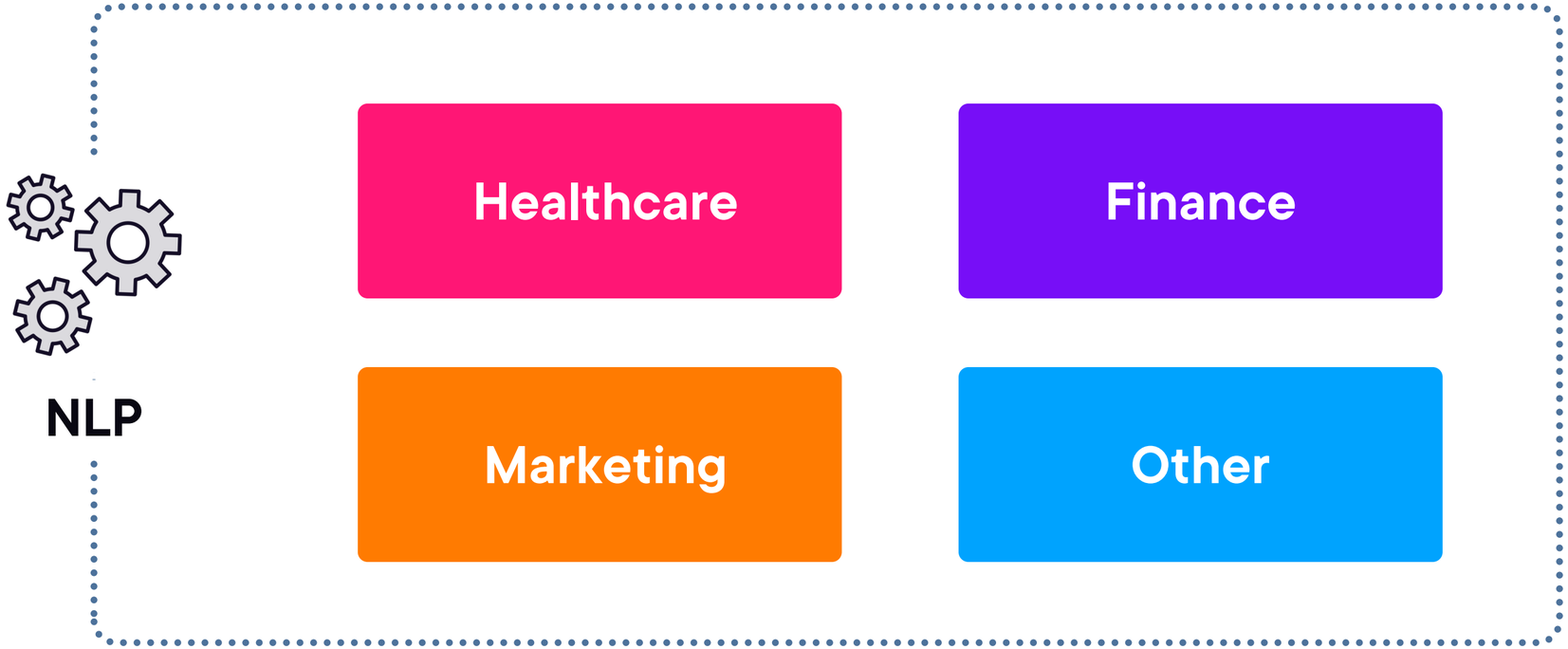
Language translation

Spelling correction



Chat bots

Industries



Overview

NLP

- Applications
- History
- Usage

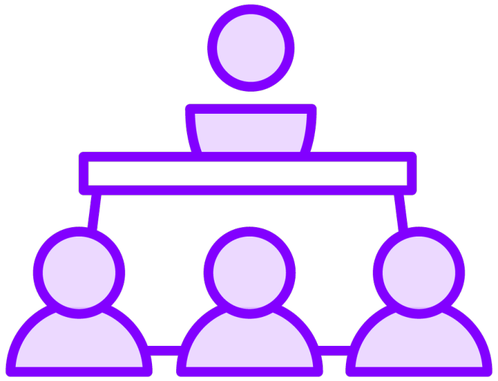
Natural language

Hands-on learning

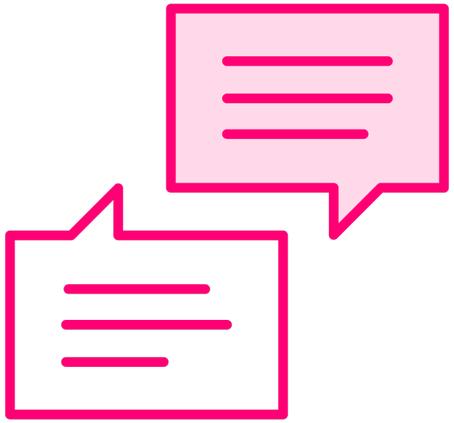
- Jupyter Notebook

What Is NLP?

Natural Language

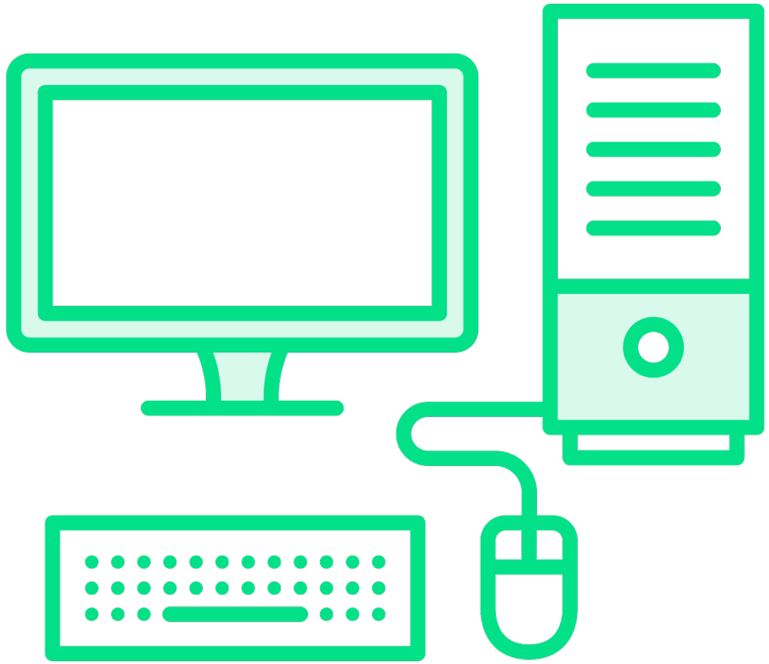


Spoken words



Text data

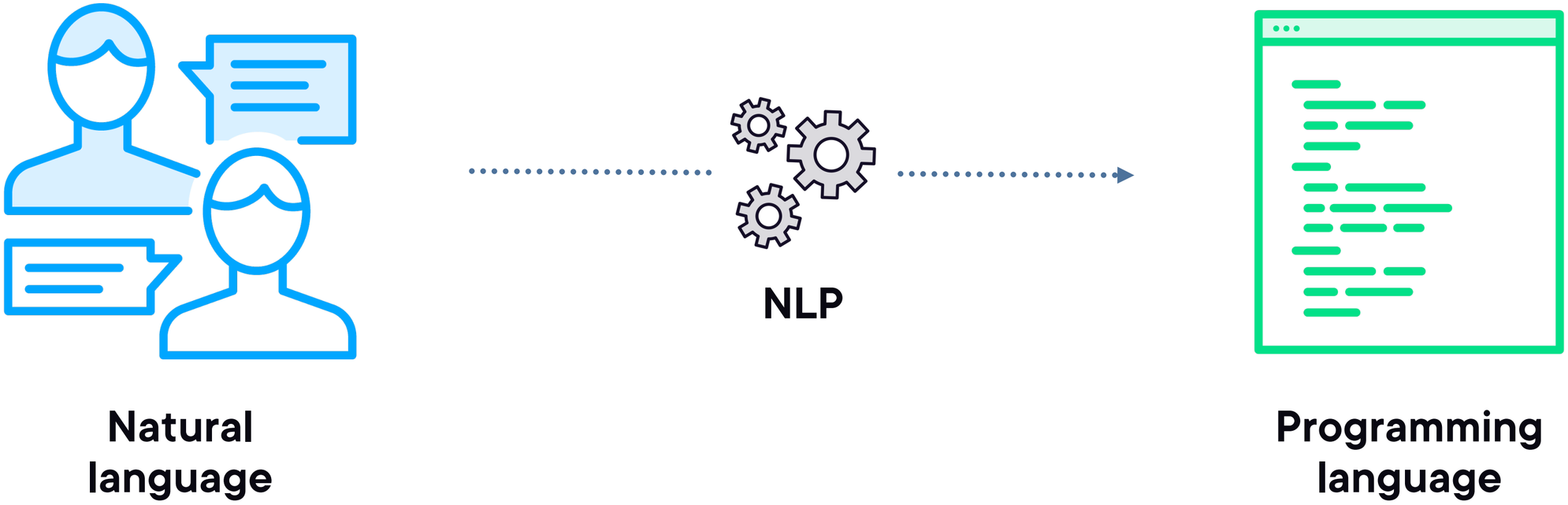
Non-verbal cues

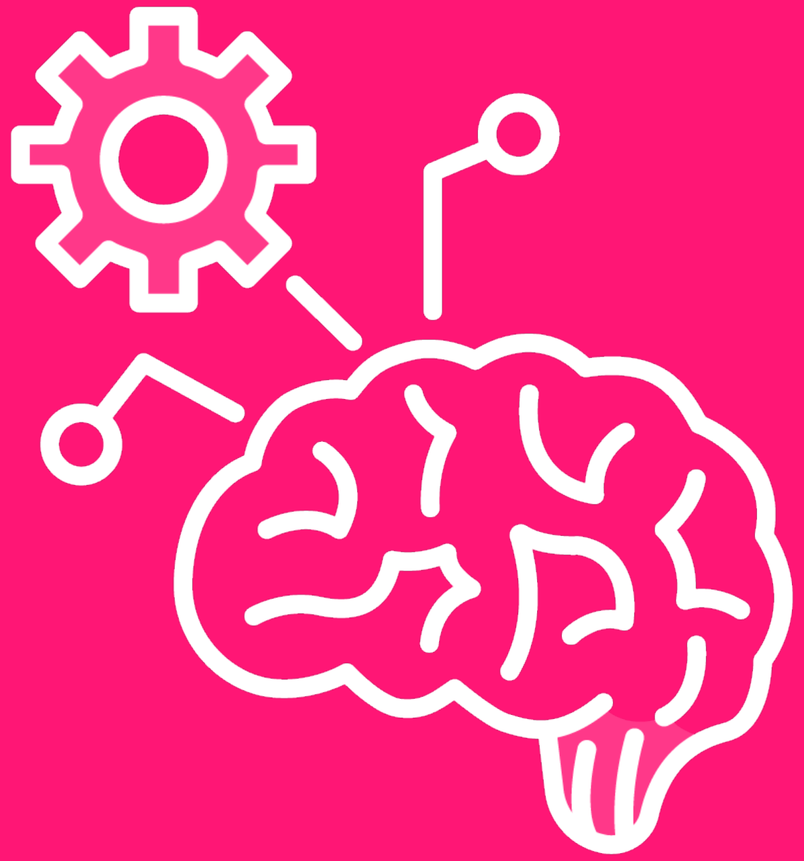


Computers don’t understand:

- Context
- Irony
- Emotion

They think in binaries


Natural language
NLP
Programming language



Natural Language Processing

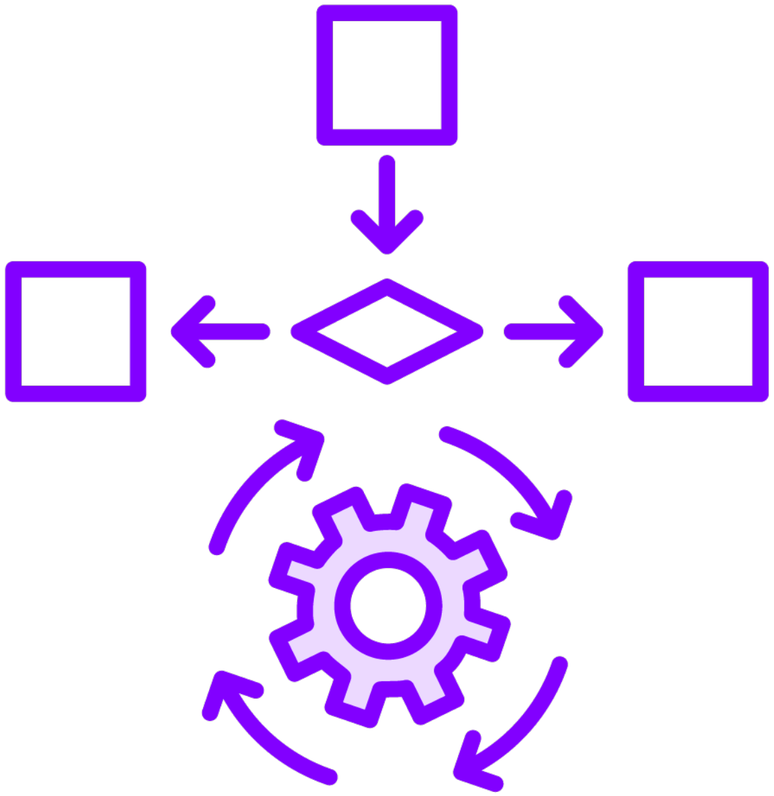
represents an automated way to understand, interpret, and generate human-like text or natural language by applying algorithms.

NLP

Linguistics

Computer science

Machine learning

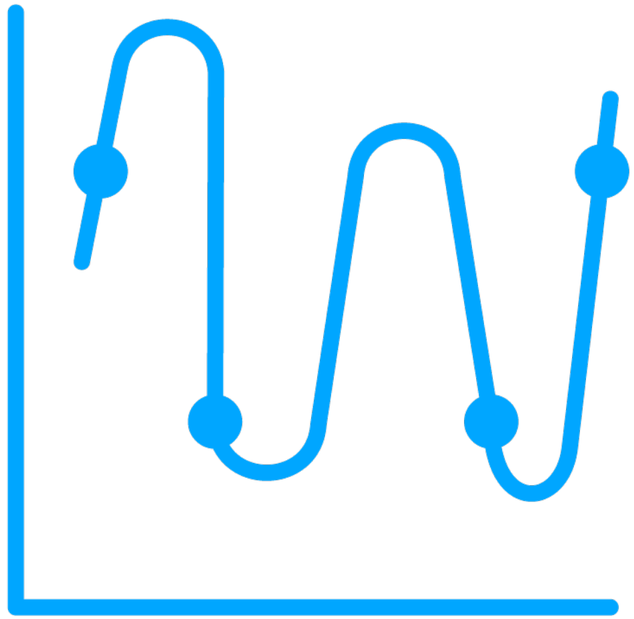


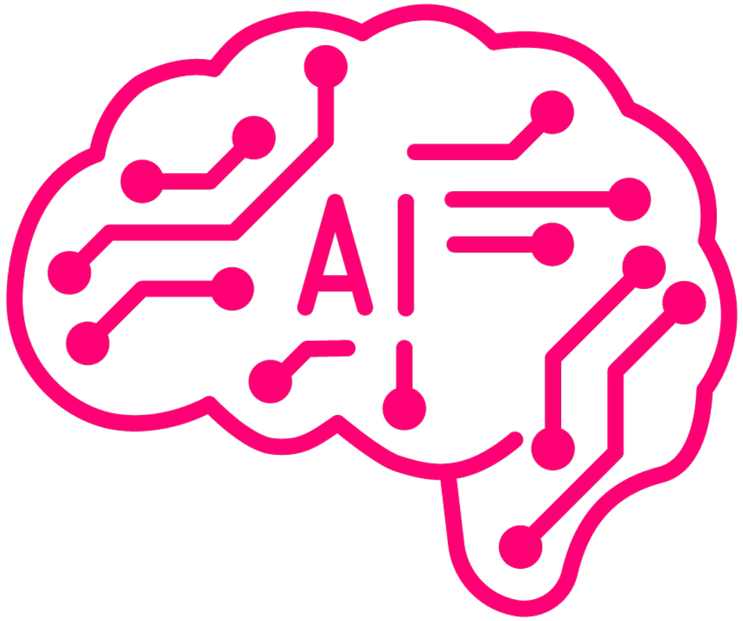
Rule-based systems:

- Appeared in 1940s
- Translation tasks

Statistical methods and machine learning:

- Gained popularity in the 1990s
- Computers learnt from examples





Deep learning:

- Recurrent neural networks (RNNs)
- Long short-term memory networks (LSTMs)
- Transformer models

Transformers

Were introduced in the 2017 paper "Attention is all you need" by Vaswani et al.

Process language using parallel token analysis

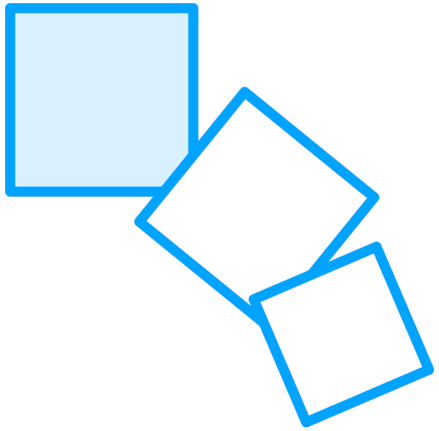
Enable the training of Large Language Models (LLMs)

Transformers

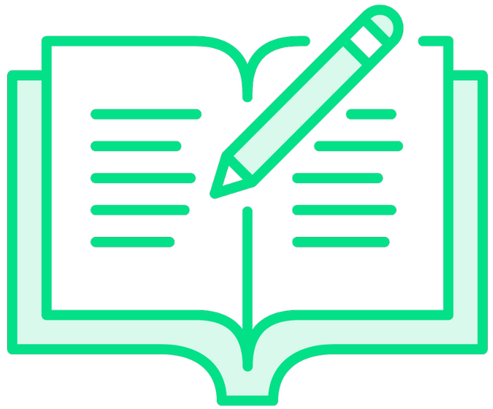
BERT (Bidirectional Encoder Representations from Transformers)

GPT (Generative Pretrained Transformer)

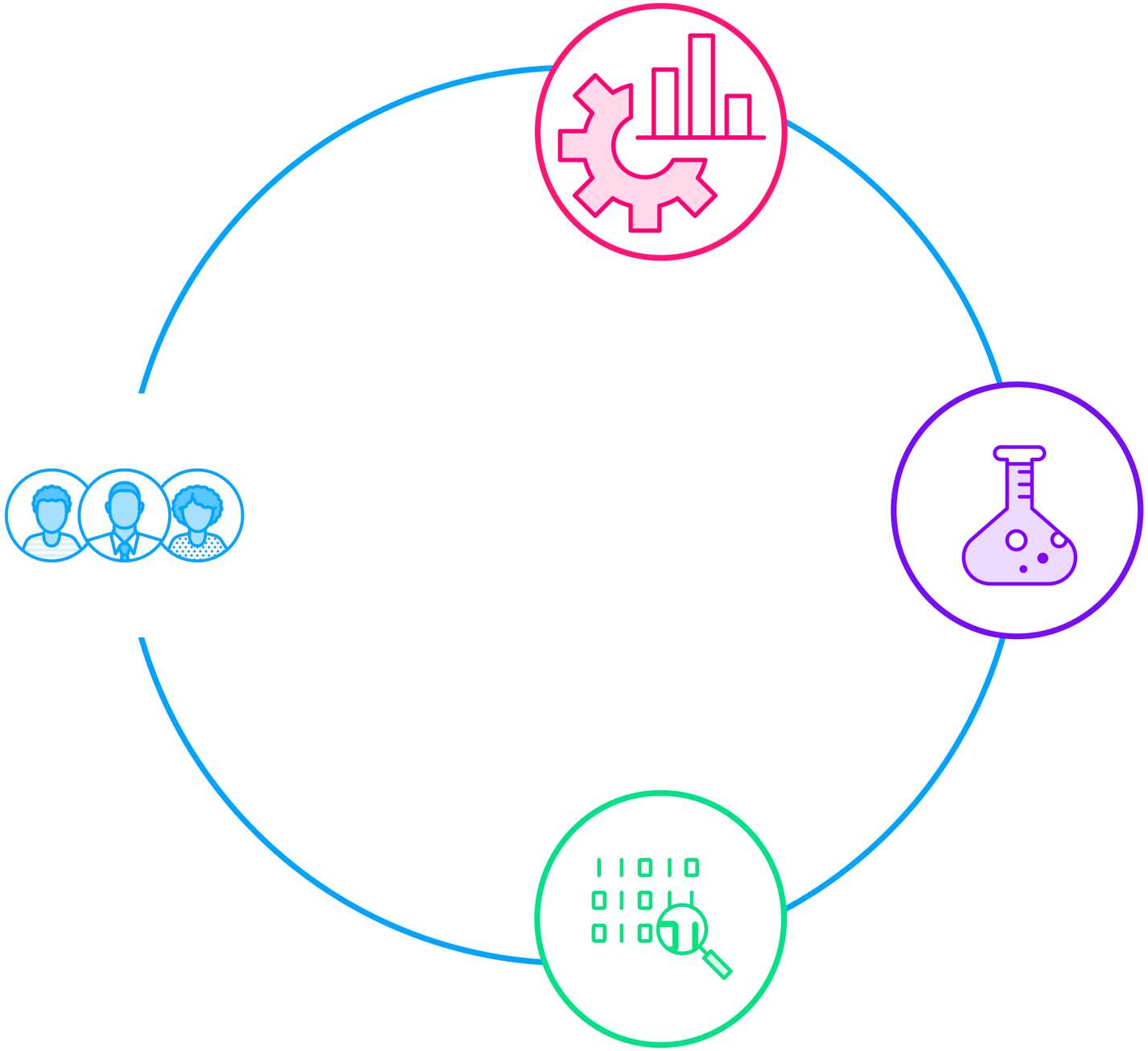
NLP Today



Dynamic

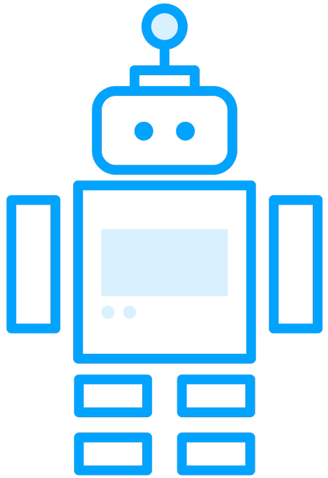


Ongoing research

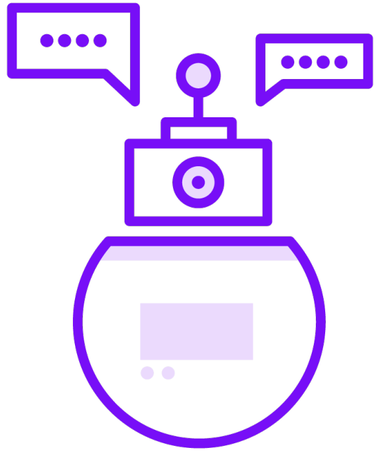


NLP Applications

Information extraction



Task automation



Enhance user experience

Industries

Healthcare

Finance

Retail

Manufacturing

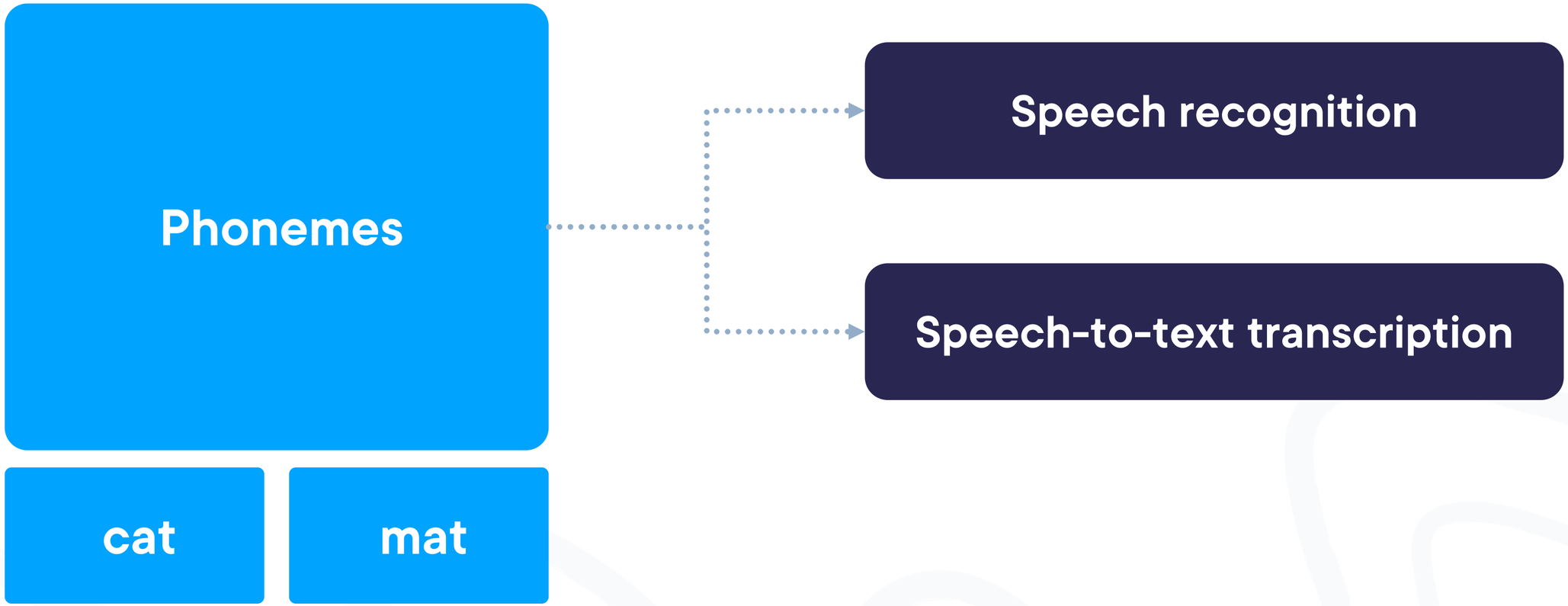
Other

Natural Language Terminology

Morphology

The study of word formation

Morphology


Phonemes
cat
mat
Speech recognition
Speech-to-text transcription

Morphology

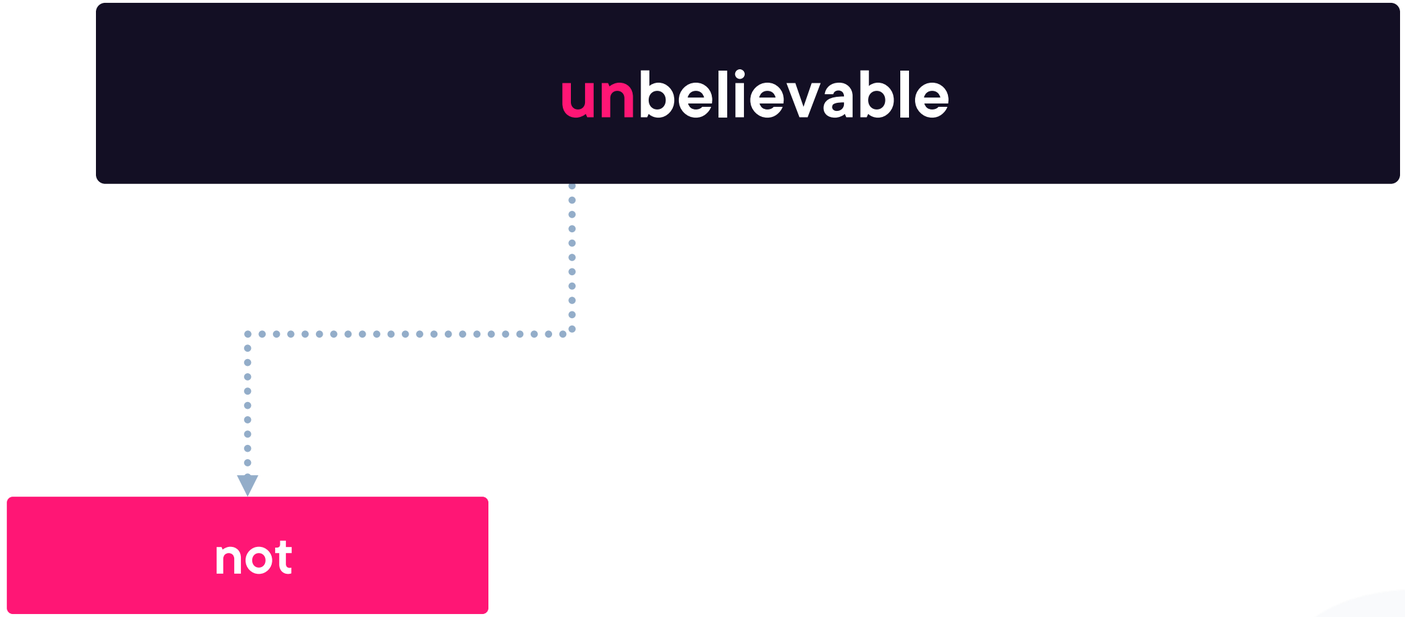
Phonemes

Morphemes

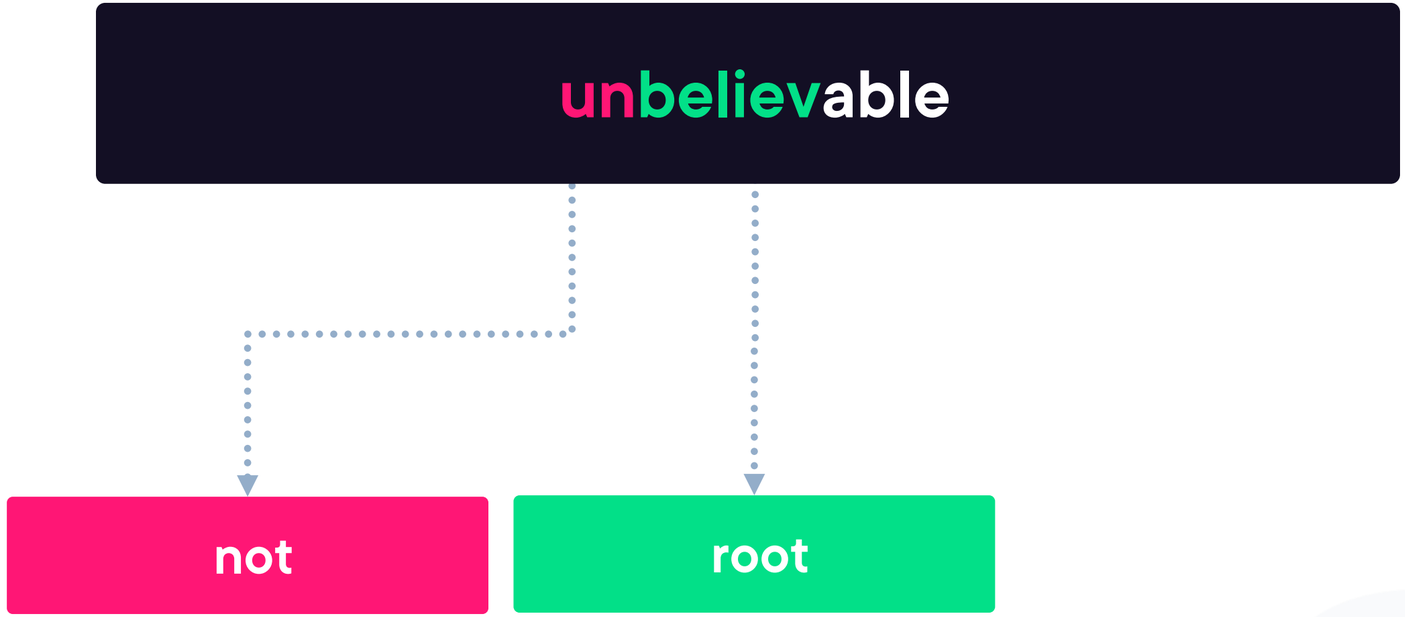
Morphemes

unbelievable

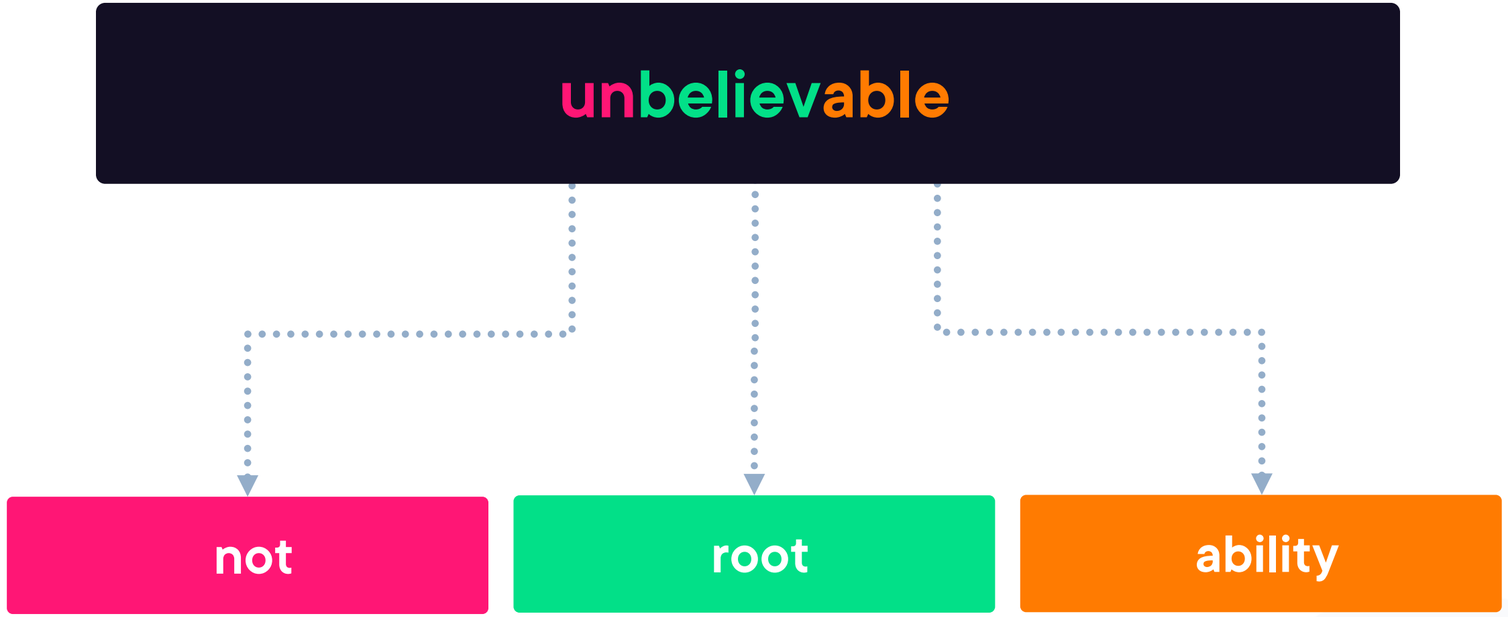
Morphemes



Morphemes



Morphemes



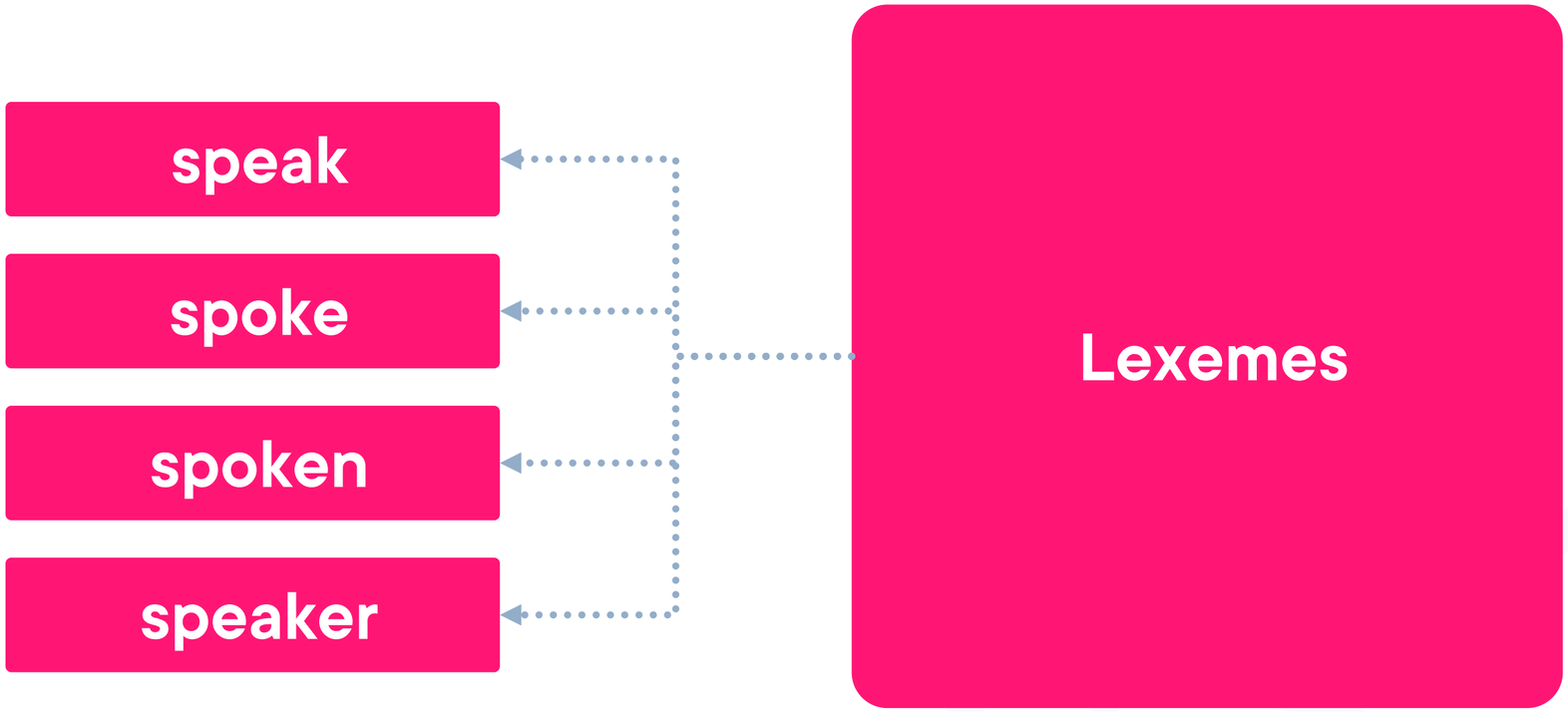
Morphology

Phonemes

Morpheme

Lexemes

Morphology

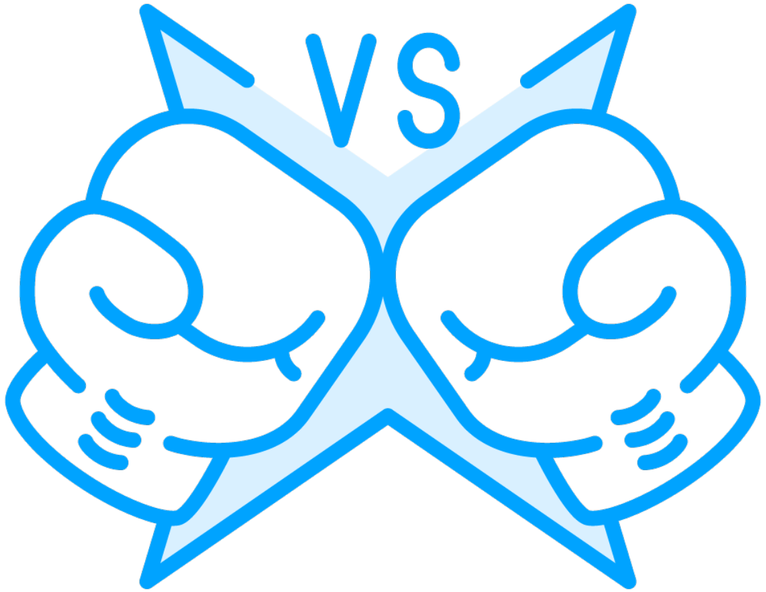


Stemming

Lemmatization

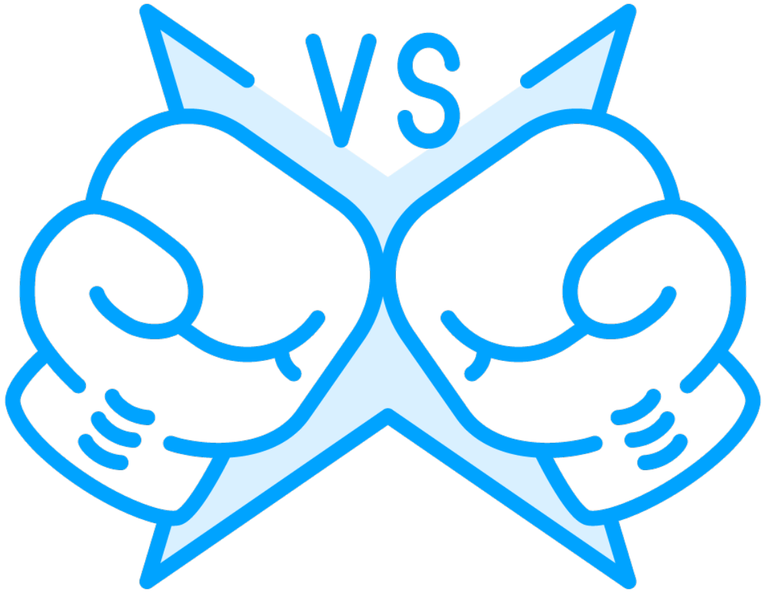
Syntax

The set of rules organizing words into sentences



I have a cat and a dog.

Cat I a have dog and.



I have a cat and a dog.

Cat I a have dog and.

Computers parse and analyze each word's role within a sentence, using POS tagging and statistical models.

Syntax

Enables machines to:

- Process our requests
- Correct our writing
- Translate languages

Semantics

Refers to the meaning of words and sentences within context.

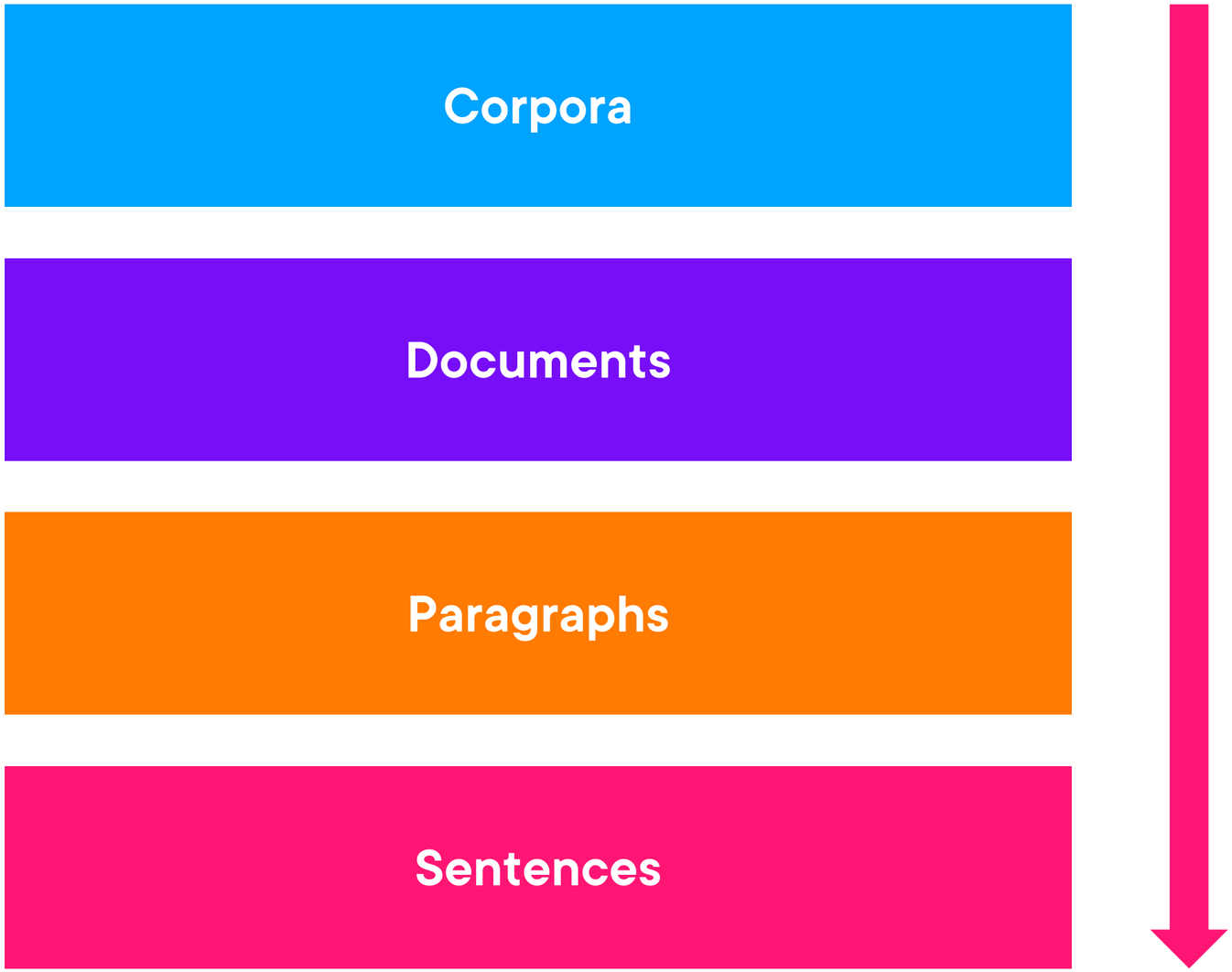
Semantics

RNNs

Transformers

I am feeling blue.

Text Structure



NLP Challenges

Intonation

Rhythm

NLP Toolkit

Toolkit

NLTK

spaCy

CoreNLP

KerasNLP

TensorFlow Text

Hugging Face

Demo

Install libraries

Text operations

Summary

NLP

NLP’s evolution

Linguistic fundamentals

Text operations

Up Next:

Exploring the NLP Pipeline