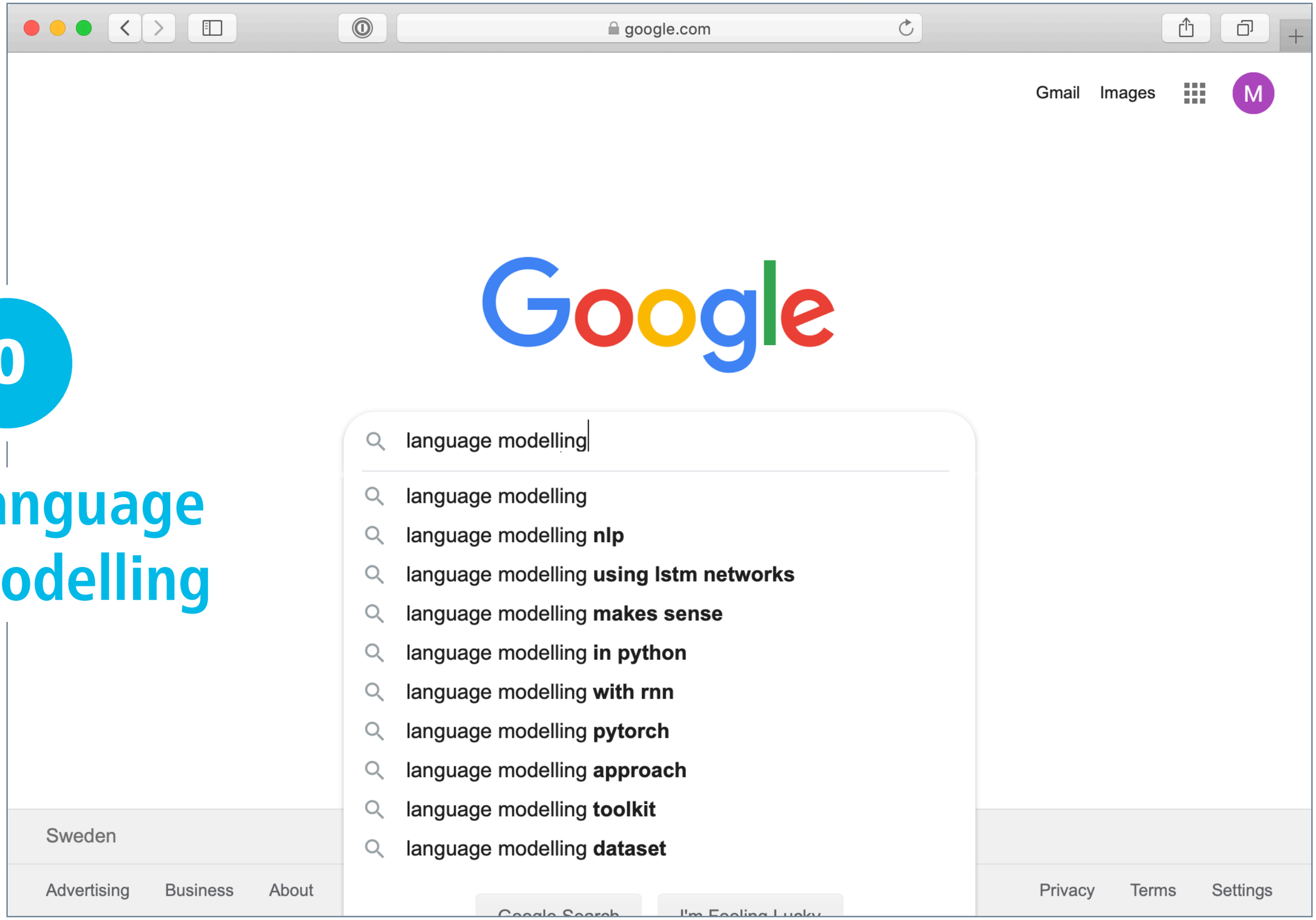
Natural Language Processing

Course overview

Marco Kuhlmann
Department of Computer and Information Science

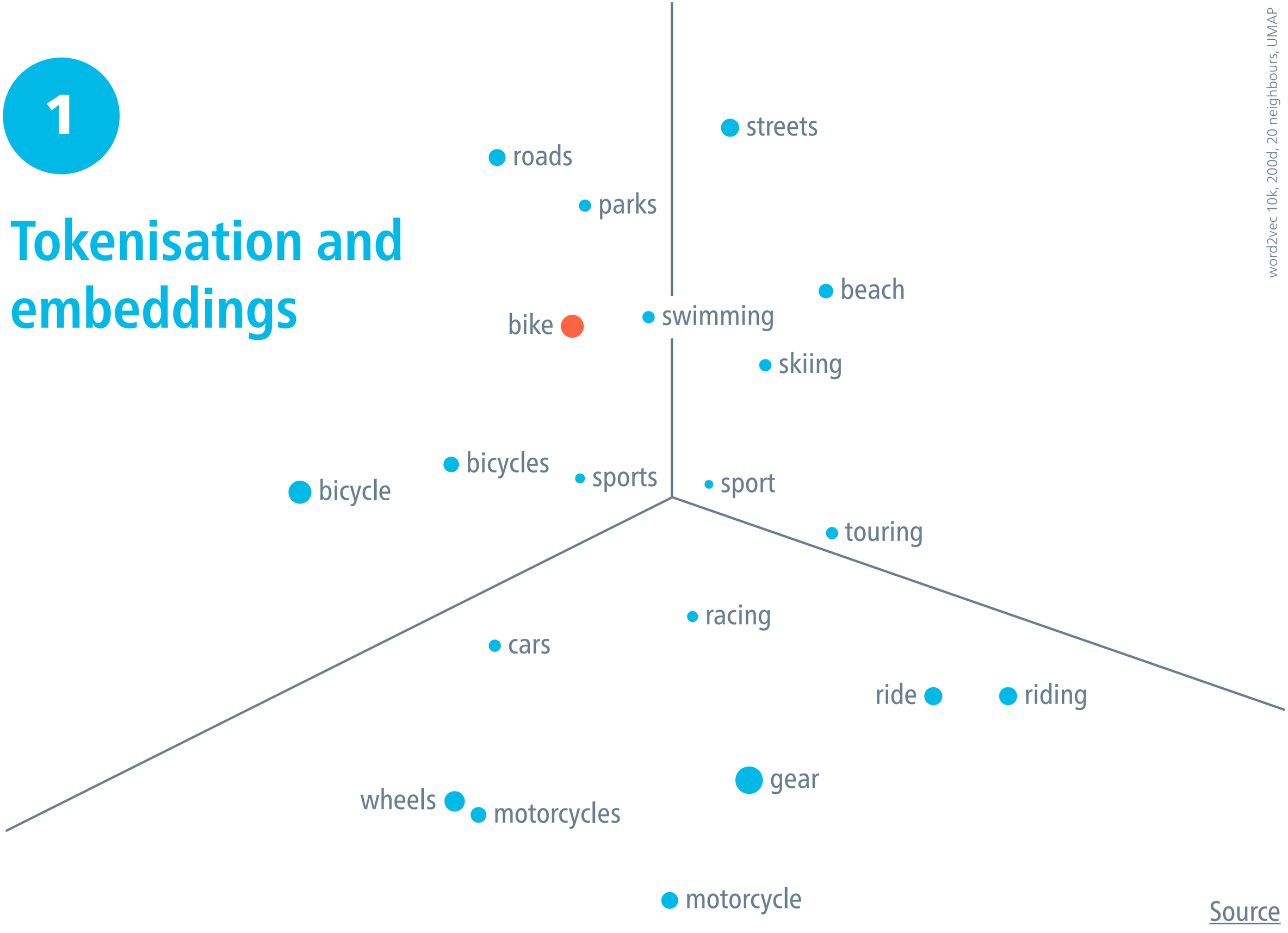
0

Language modelling



1

Tokenisation and embeddings

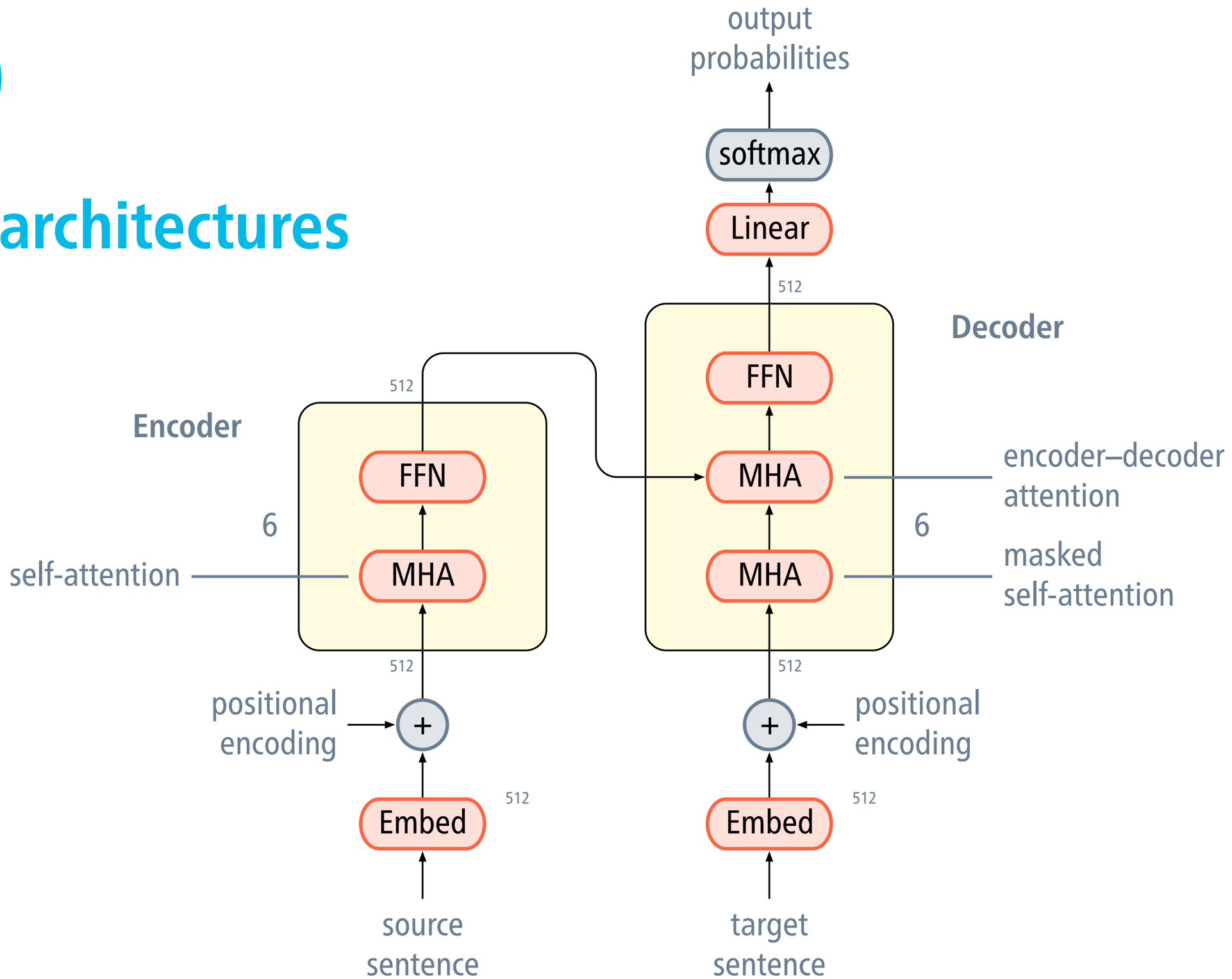


word2vec 10k, 200d, 20 neighbours, UMAP

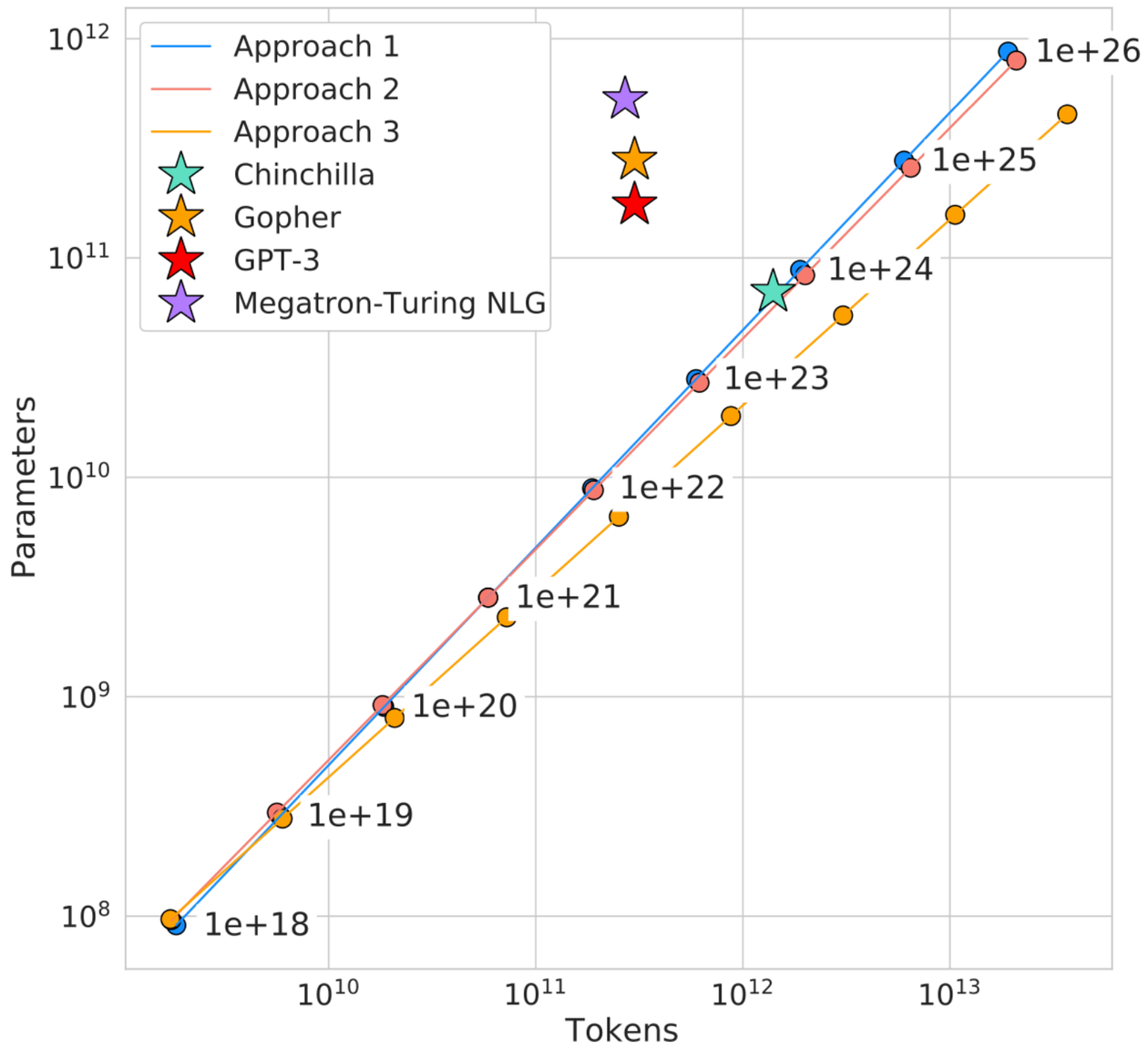
Source

2

LLM architectures



3 Pre-training



Hoffmann et al. (2022)

4 Alignment and evaluation

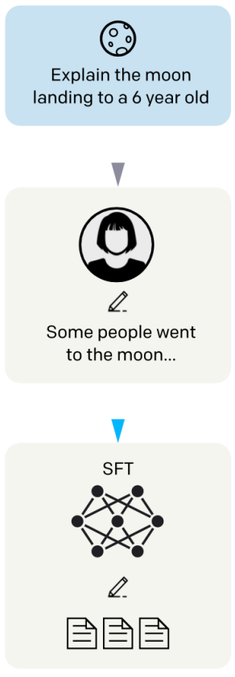
Step 1

Collect demonstration data, and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3 with supervised learning.



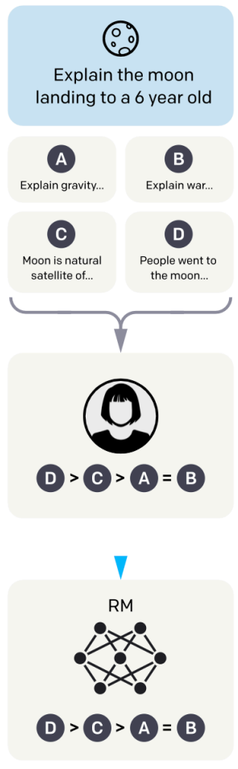
Step 2

Collect comparison data, and train a reward model.

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.



Step 3

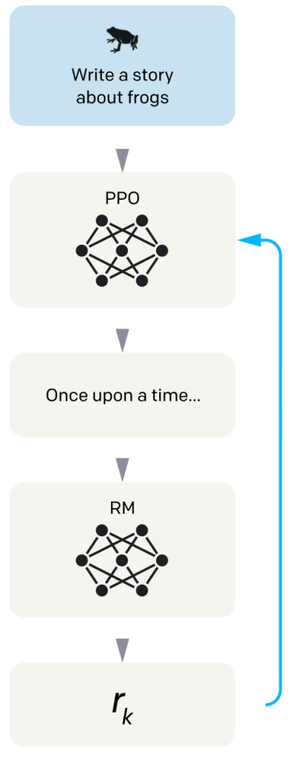
Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



[Ouyang et al. (2022)]