




ChatGPT: the revolution is coming?

KYOSS
Liang Yan

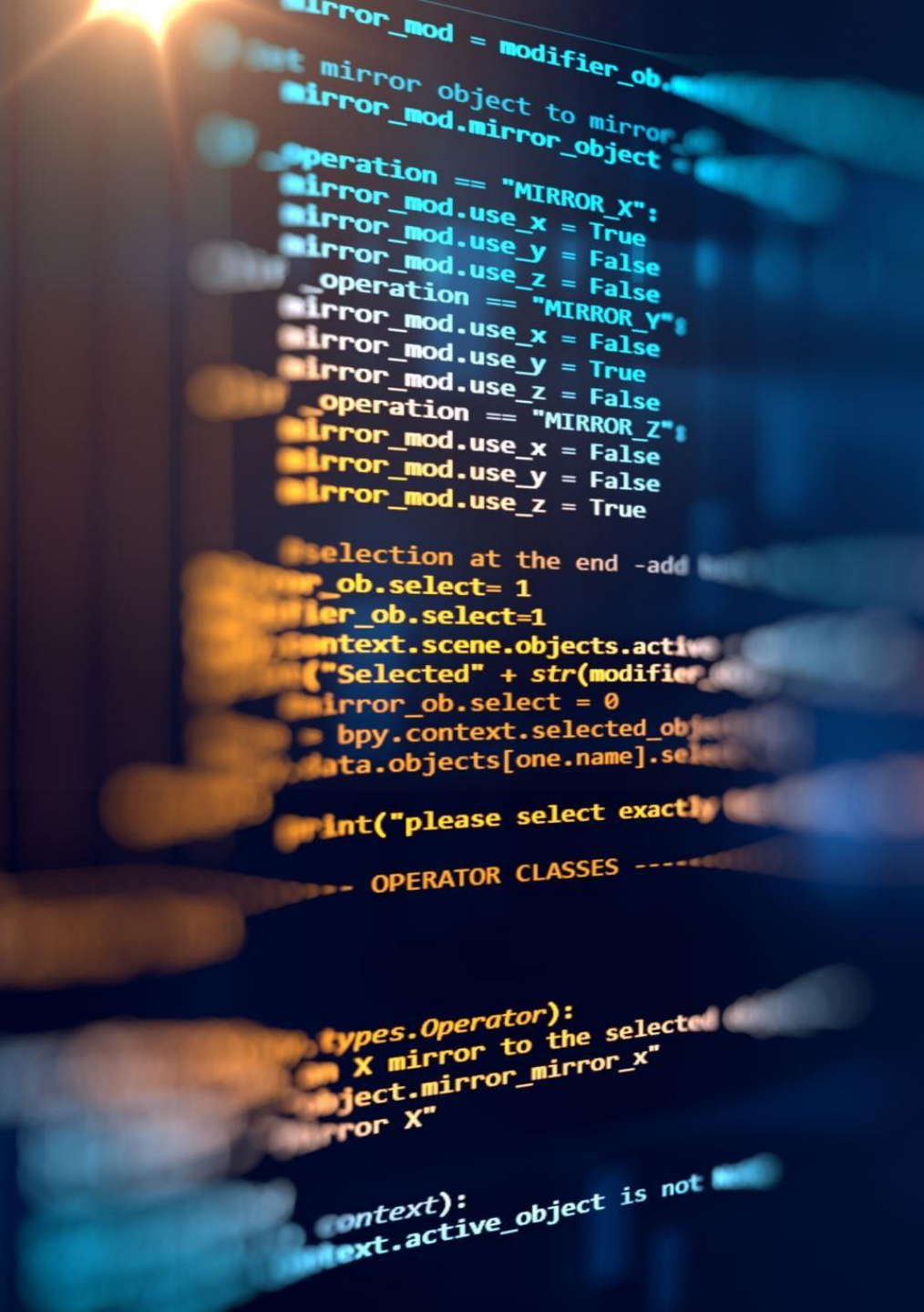


ChatGPT

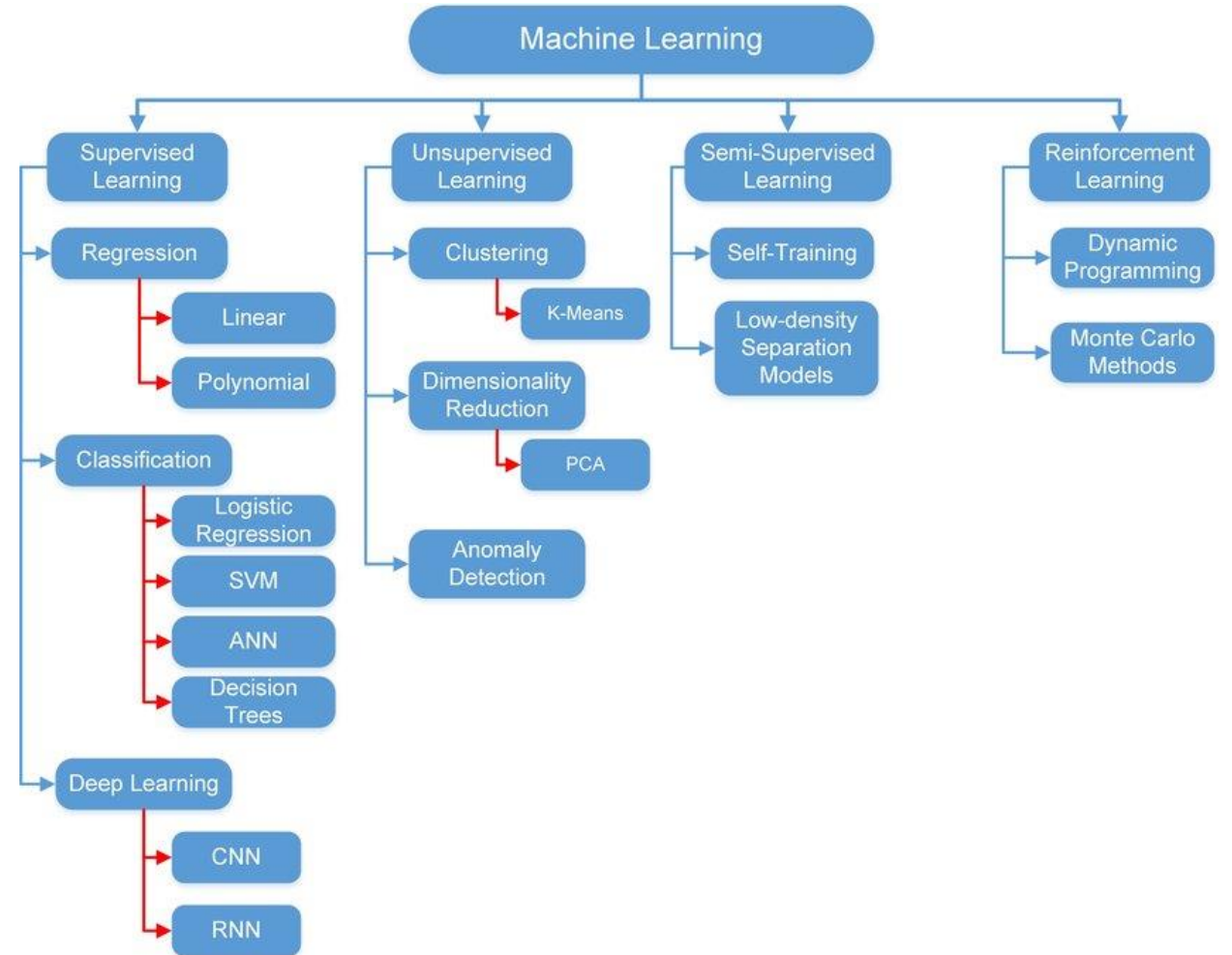
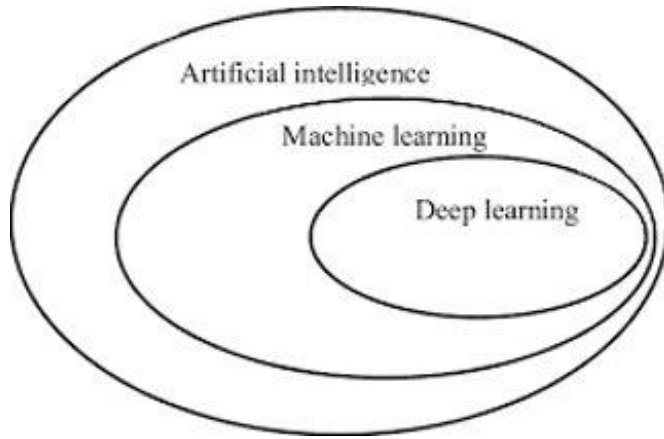
- A chatbot from OpenAI
 - AIGC: AI-Generated Content
 - AI painting
 - AI chat
 - AGI: Artificial general intelligence
- 

ChatGPT

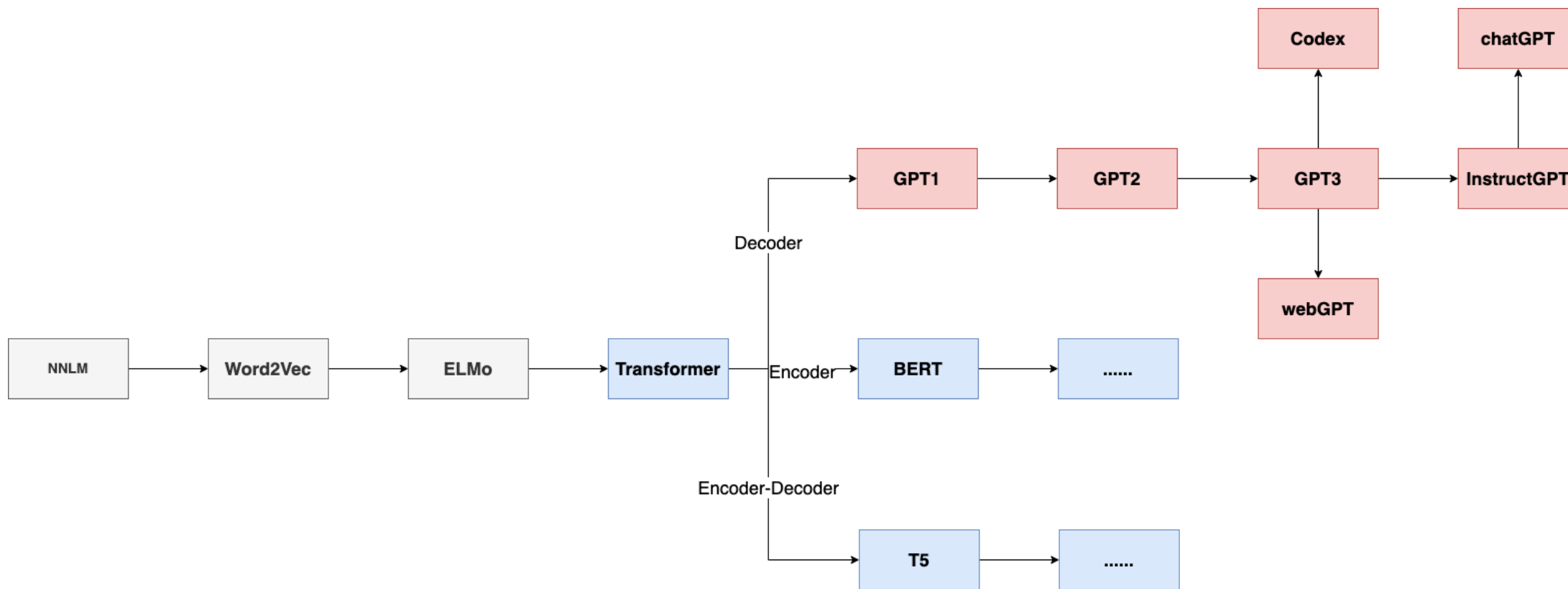
- Chat => Language Model
NLP(Nature Language Processing)
- GPT => GPT-3.5



Artificial Intelligence



Language Model



Ref: https://blog.csdn.net/2201_75869398/article/details/128337270

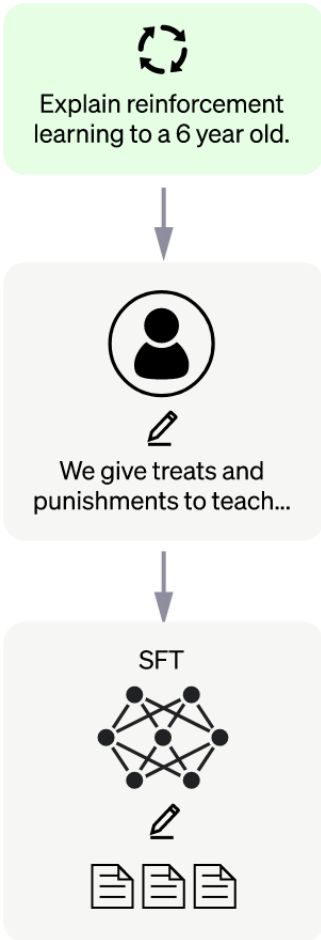
ChatGPT

- Tune pre-trained GPT 3.5 model
- Train a reward model
- Reinforcement Learning with reward model

Step 1

Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.



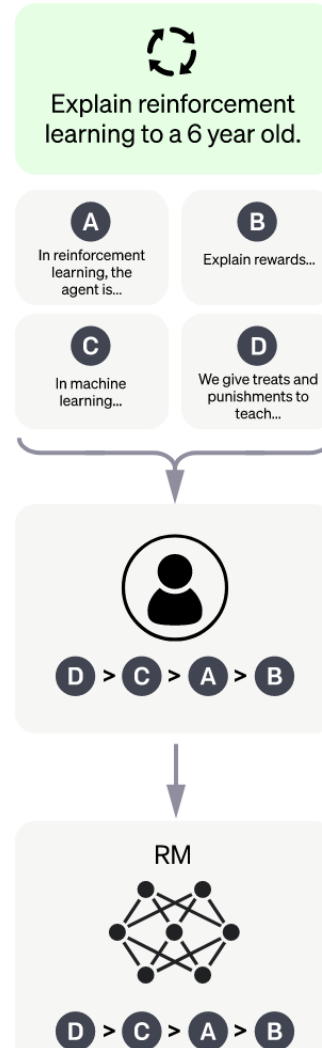
A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.

Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.



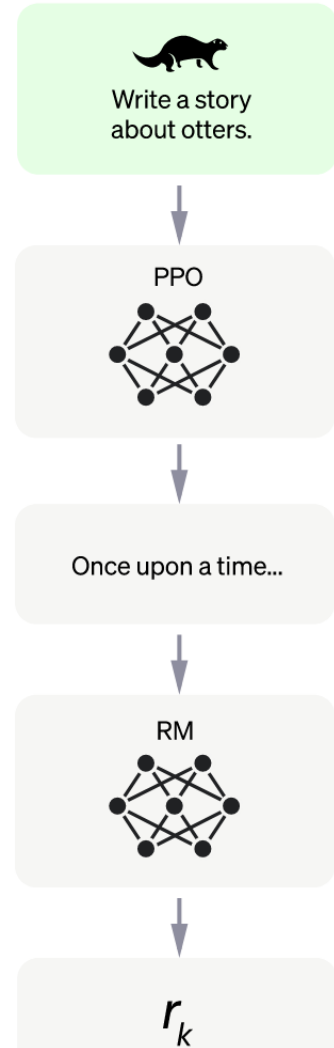
A labeler ranks the outputs from best to worst.

This data is used to train our reward model.

Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

A new prompt is sampled from the dataset.



The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



Q & A

Thanks!

