Parameter Efficient Fine-Tuning

Advanced NLP: Summer 2023

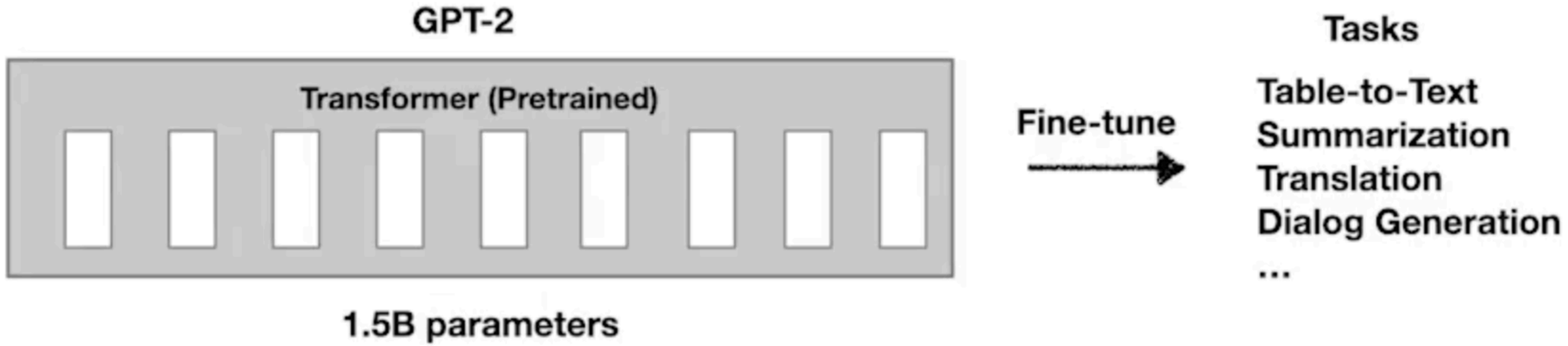
Anoop Sarkar

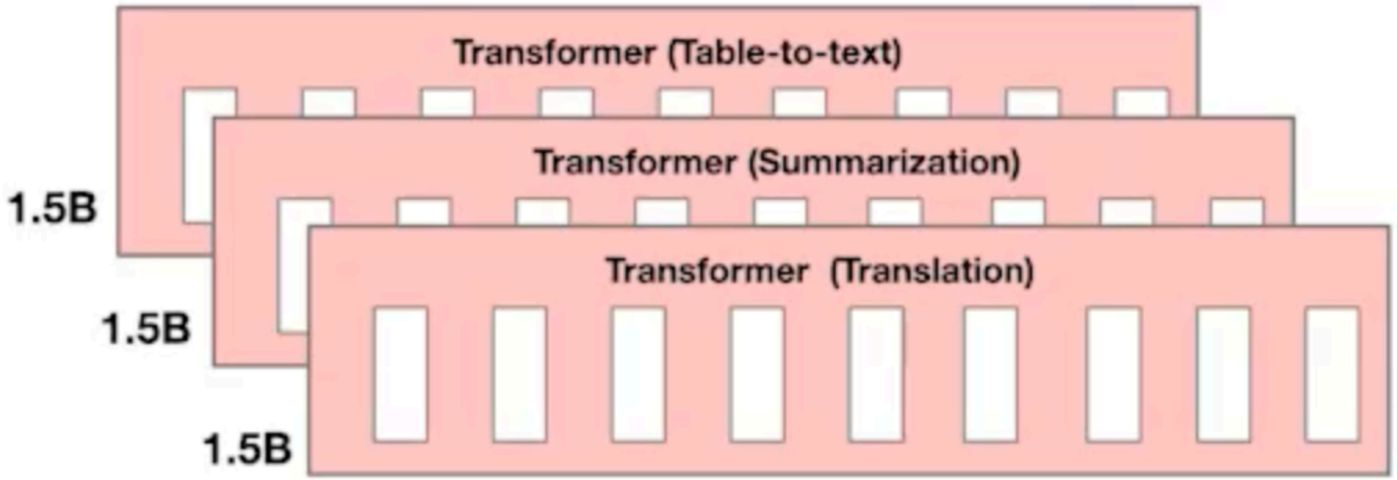
Prefix Tuning

Li and Liang, ACL 2021

https://aclanthology.org/2021.acl-long.353

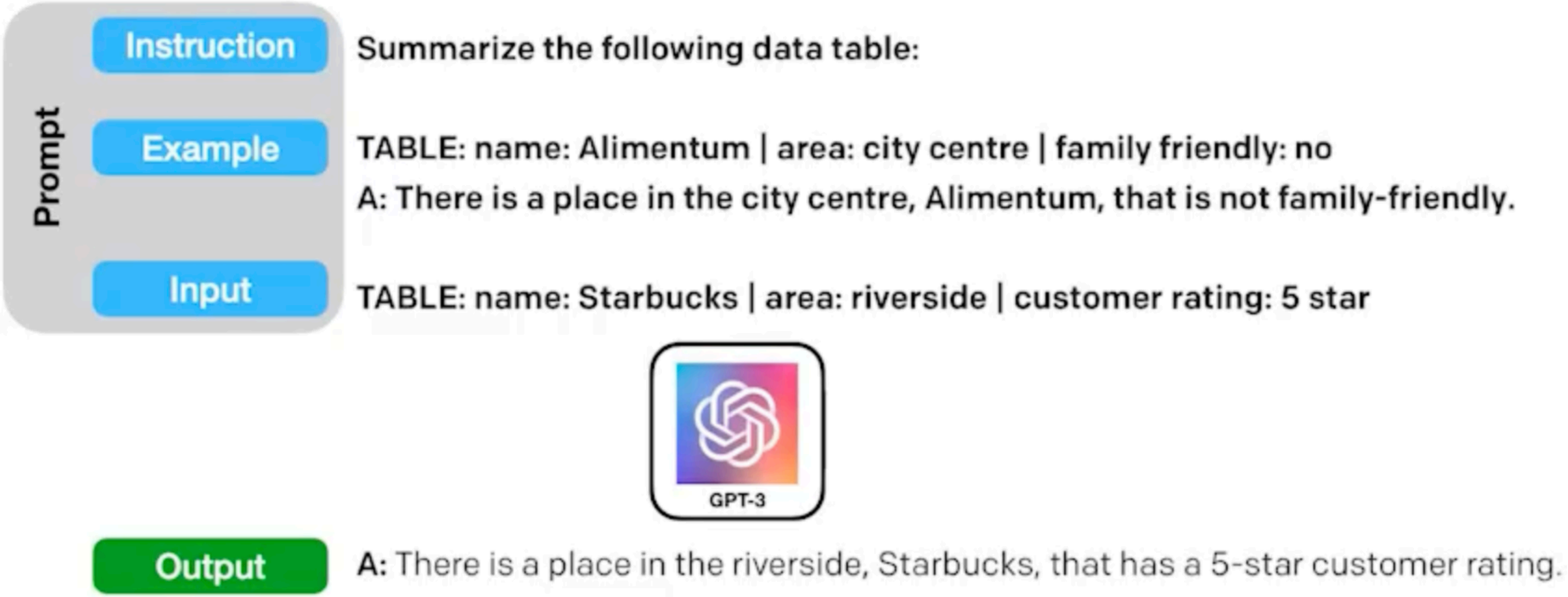
Why not just use fine-tuning





Each task requires a full model copy

In-context learning using prompts



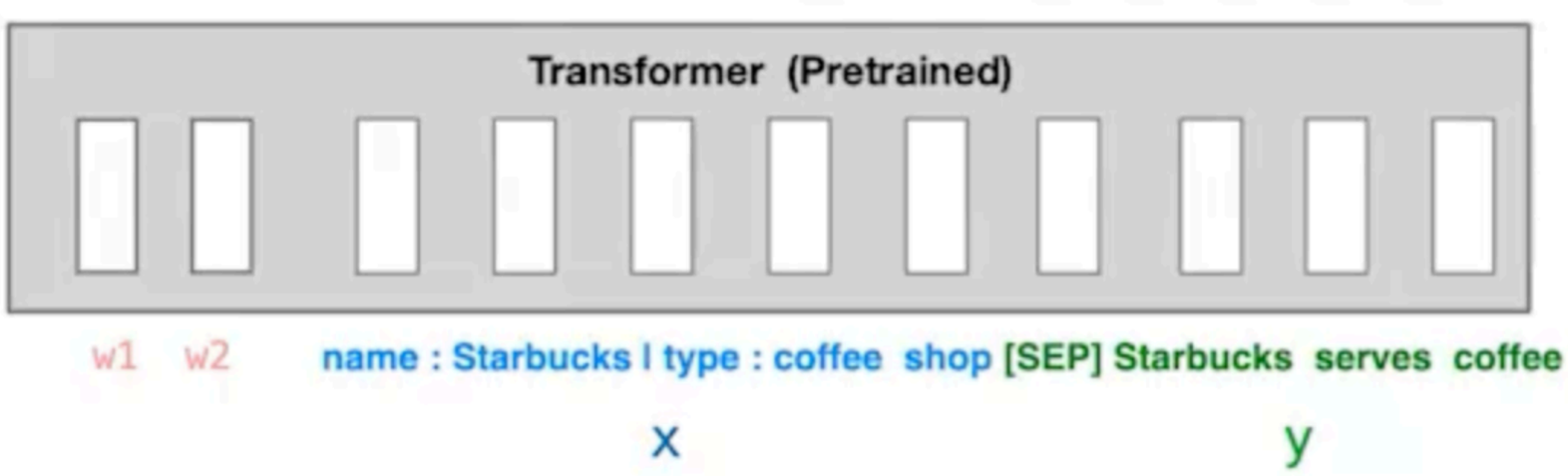
- No task specific fine-tuning
- Preserves the LM

- Cannot use large training set
- Manual prompts can be suboptimal
- Cannot be used with smaller LMs like GPT-2

Prefix Tuning

Intuition

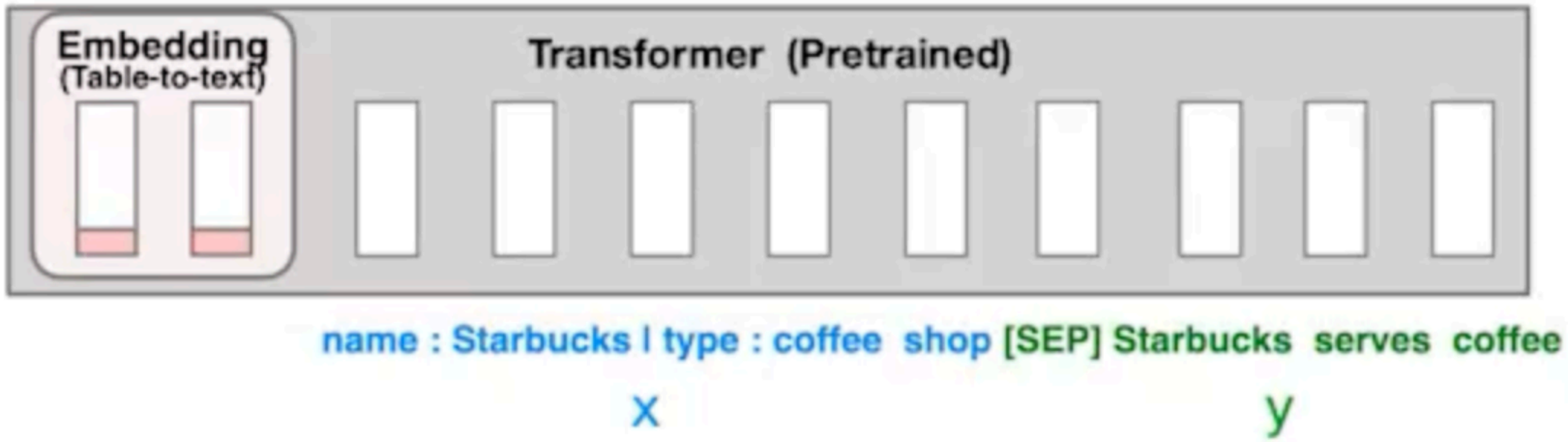
- Learn a good instruction that can steer the LM to produce the right output
- Optimize finding actual words
- Involves discrete optimization which is challenging and not expressive



Prefix Tuning

Intuition

- Optimize the instruction as continuous word embeddings
- More expressive
- Limits the scope of the prompt to a input embeddings



Prefix Tuning

Intuition

- Optimize the instruction as prefix activation for all layers in the instruction
- Very expressive
- All the layers of the prefix can be tuned to create the most expressive prompt