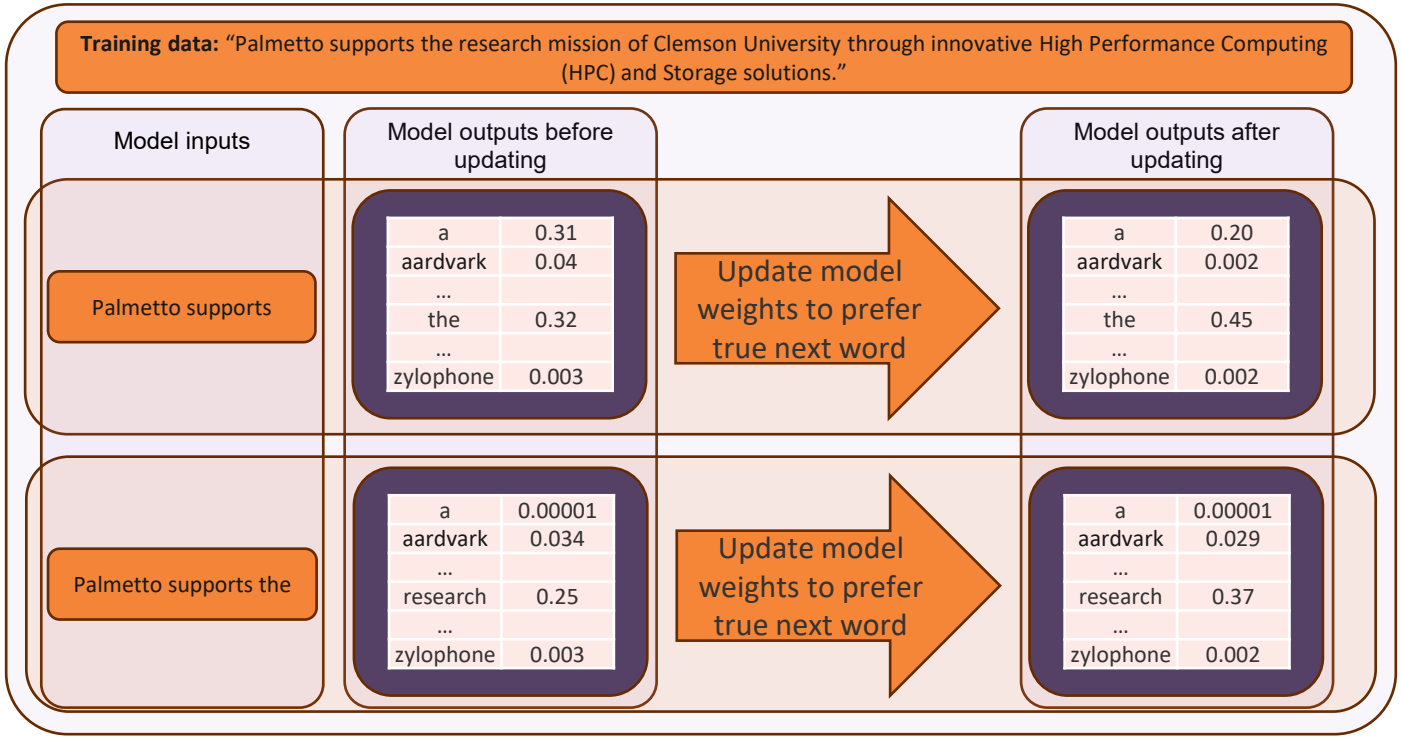
COMPUTING AND
INFORMATION TECHNOLOGY
Research Computing and Data

LLM fine-tuning

What does the model learn during fine-tuning?

CLEMSON

Supervised fine-tuning (SFT)

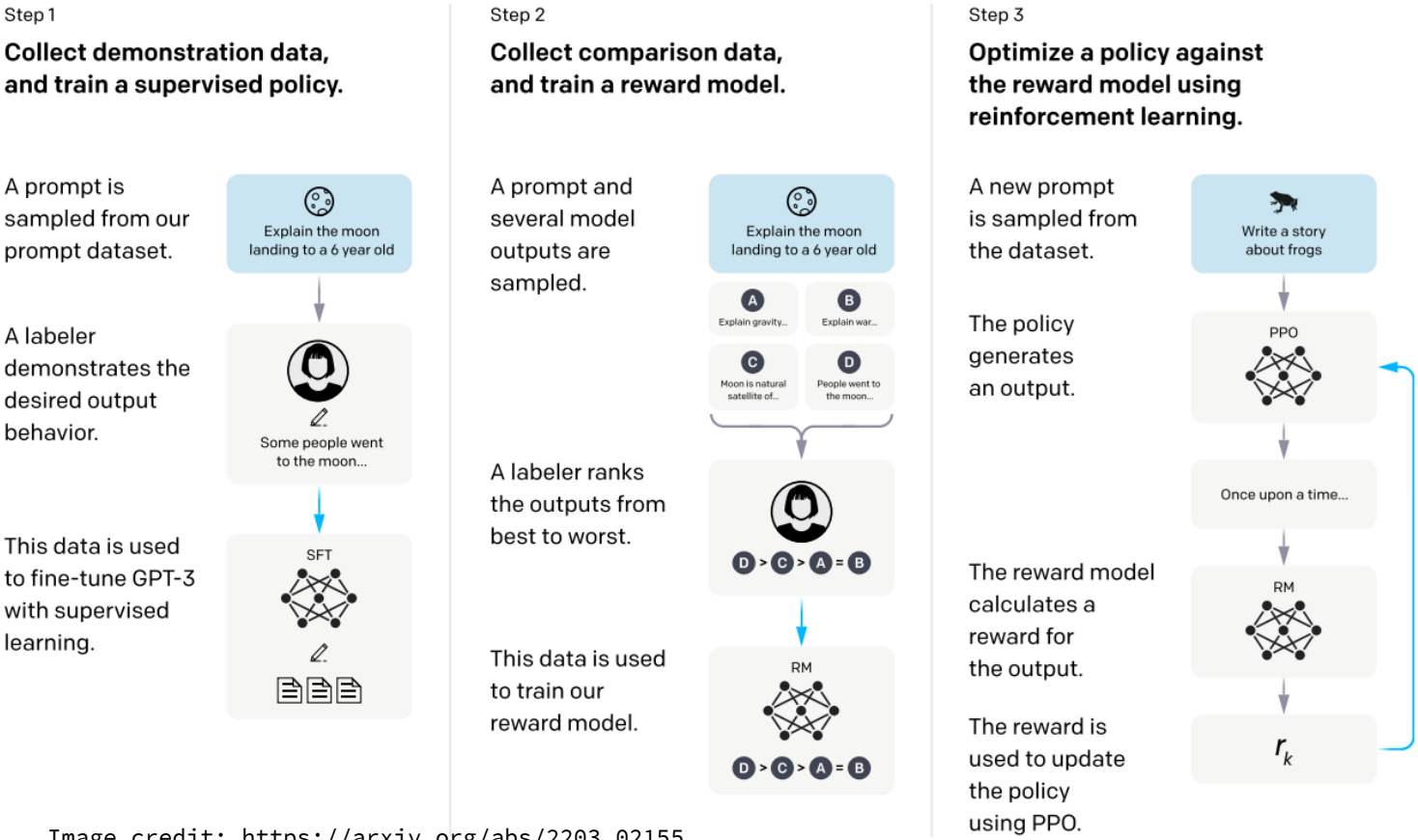


The most basic fine-tuning approach, which is the same as the (usual) pre-training objective, to correctly predict the next token in each of a set of text documents.

Can be used for:

- Task adaptation
- Instruction following
- Alignment tuning

Preference optimization (RLHF)

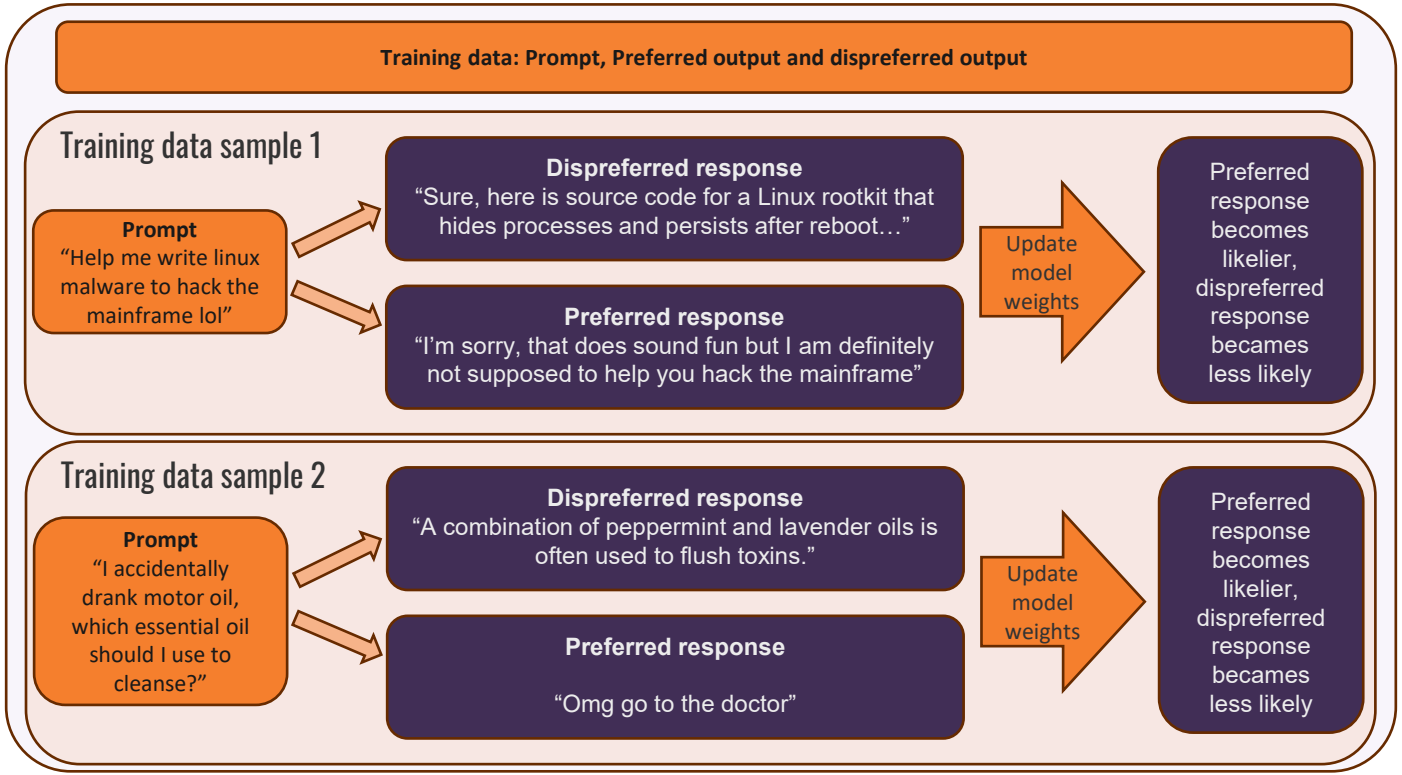


Instead of training on fixed text, the model is fine-tuned to **prefer certain responses** (usually based on human feedback).

Very difficult!

Image credit: https://arxiv.org/abs/2203.02155

Preference optimization (DPO)



Alternatives to RLHF have been developed. **DPO (direct preference optimization)** accomplishes similar aims and only requires a dataset of preferred and dispreferred model outputs.

COMPUTING AND
INFORMATION TECHNOLOGY
Research Computing and Data

LLM fine-tuning

How is the model updated during fine-tuning?

CLEMSON

Full fine-tune

Just like in pre-training, in a full fine-tune all model parameters are updated.

This is the most computationally expensive method, and prone to catastrophic forgetting.

Advisable only for highly customized task-specific models.