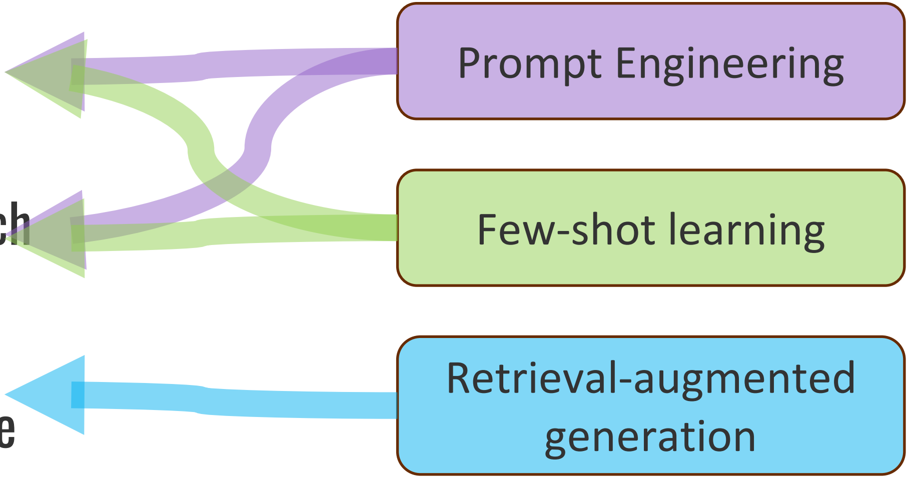
Why people want to fine-tune (but don’t need to)

- The task the LLM will perform is not present in the training data
- The task the LLM will perform is much more specific and narrow than the training data
- To provide additional knowledge base



Why you should fine-tune anyway

- Fine-tuning can achieve (sometimes small) quality improvements over other methods
- In some cases, your FSL examples might be too large for a model's context window
- Smaller LLMs can benefit more from fine-tuning than large ones
- If you expect to use the model for inference many times, fine-tuning can reduce your prompt size and thus reduce the cost of inference