DeepLearning.AI

(An Incomplete List of) Popular LLM Evals

Human Preferences for chat	**Chatbot Arena**
LLM as a judge for chat	Alpaca Eval MT Bench **Arena Hard V1 / V2**
Static Benchmarks for Instruct LLM	**LivecodeBench** **AIME 2024 / 2025** GPQA MMLU Pro IFEval
Function Calling & Agent	BFCL V2 / V3 NexusBench V1 / V2 **TauBench** **ToolSandbox**

It's easy to improve any one of the benchmarks.

It's much harder to improve **without degrading other domains**.

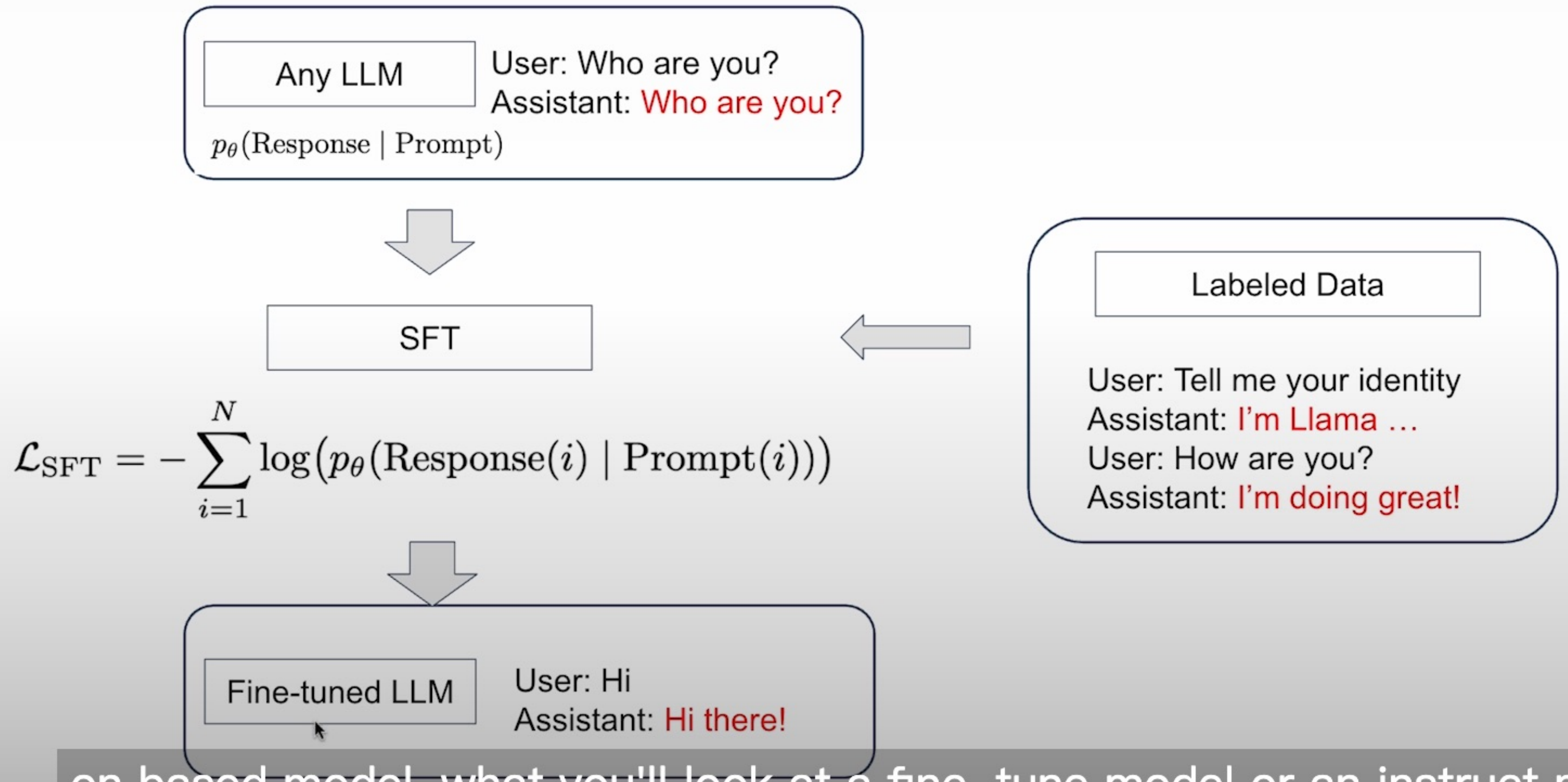
Do you really need post-training?

Use Cases	Methods	Characteristics
Follow a few instructions (do not discuss XXX)	Prompting	Simple yet brittle: models may not always follow all instructions
Query real-time database or knowledgebase	Retrieval- Augmented Generation (RAG) or Search	Adapt to rapidly-changing knowledgebase
Create a medical LLM / Cybersecurity LLM	Continual Pre-training + Post-training	Inject large-scale domain knowledge (>1B tokens) not seen during pre-training
Follow 20+ instructions tightly; Improve targeted capabilities ("Create a strong SQL / function calling / reasoning model")	Post-training	Reliably change model behavior & improve targeted capabilities; May degrade other capabilities if not done right

how to do post-training and when to do post-training.

如何進行後訓練以及何時進行後訓練

SFT: Imitating Example Responses



on based model, what you'll look at a fine-tune model or an instruct model,
基於基礎模型，你將會看到一個微調模型或是指令模型，

Best Use Cases for SFT

- **Jumpstarting new model behavior**
  - Pre-trained models -> Instruct models
  - Non-reasoning models -> reasoning models
  - Let the model uses certain tools without providing tool descriptions in the prompt

- **Improving model capabilities**
  - Distilling capabilities for small models by training on high-quality synthetic data generated from larger models

synthetic data generated by a larger model.
這些訓練資料是由較大型模型生成的合成數據

DeepLearning.AI

Principles of SFT Data Curation

- **Common methods for high-quality SFT data curation:**
  - **Distillation**: Generate responses from a stronger and larger instruct model
  - **Best of K / rejection sampling**: Generate multiple responses from the original model, select the best among them
  - **Filtering**: start from larger scale SFT dataset, filter according to the quality of responses and diversity of the prompts

- **Quality > quantity for improving capabilities:**
  - 1,000 high-quality, diverse data > 1,000,000 mixed-quality data

will be forced to imitate such response and thus degrading the performance.
將被迫模仿此類回應，從而導致效能下降。

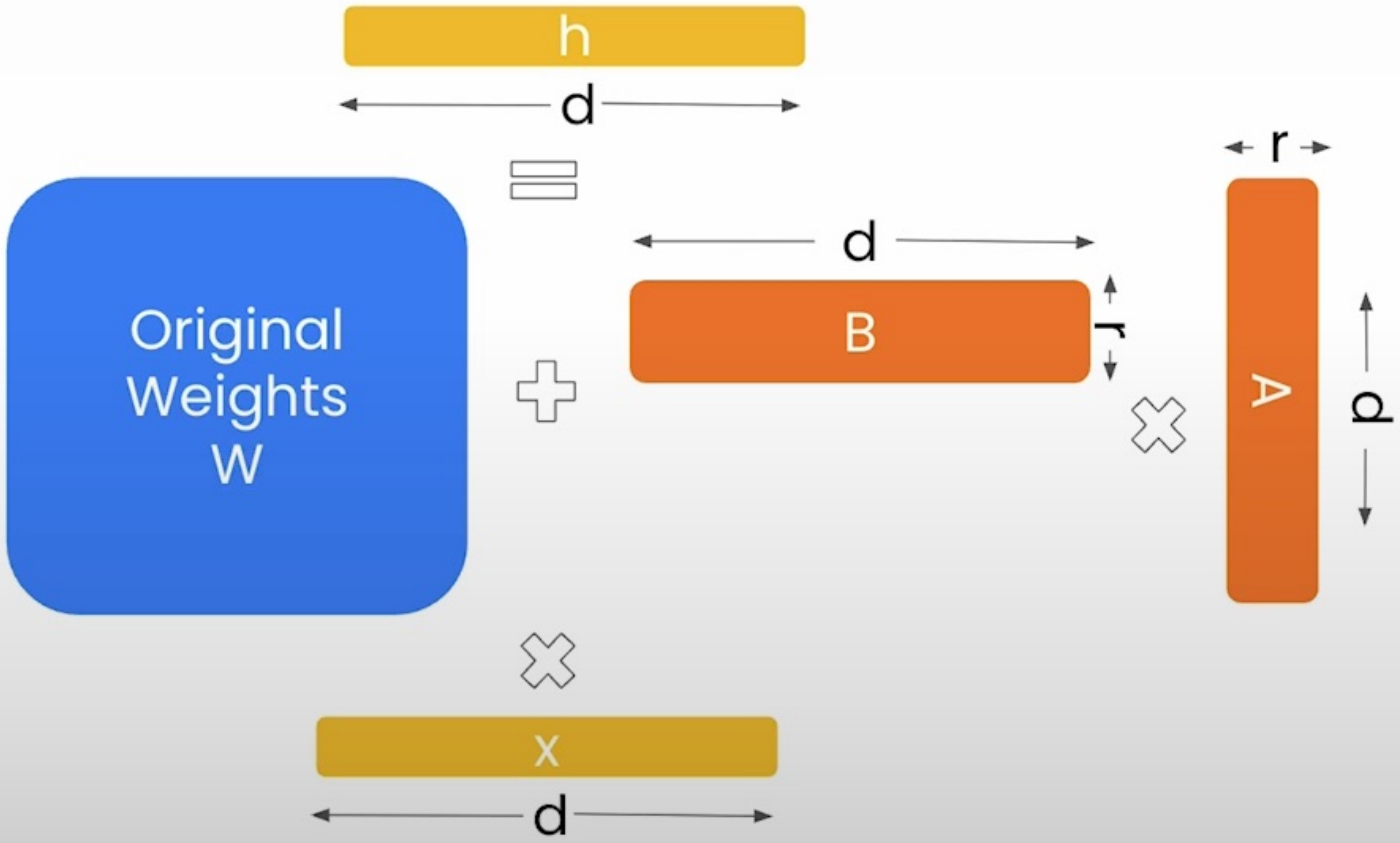
Full Fine-tuning vs Parameter Efficient Fine-tuning (PEFT)

$$h = (W + \Delta W)x$$

$$W, \Delta W \in \mathbb{R}^{d \times d}, h, x \in \mathbb{R}^{d \times 1}$$

$$h = (W + BA)x$$

$$B \in \mathbb{R}^{d \times r}, A \in \mathbb{R}^{r \times d}$$



Both full-finetuning and PEFT can be used in any of the post-training methods.
PEFT like Lora saves memory, learns less while forgets less [1]

[1] Biderman, Dan, Jacob Portes, Jose Javier Gonzalez Ortiz, Mansheej Paul, Philip Greengard, Connor Jennings, Daniel King et al. "Lora learns less and forgets less." *arXiv preprint arXiv:2405.09673* (2024).