



# Advancing the Pareto Frontier of Training Open Language Models

Mengzhou Xia  
Princeton University

# AI development is advancing fast

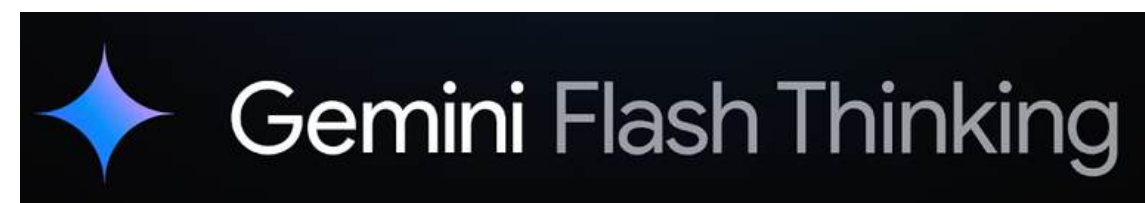
## Chat Models

2022-Now



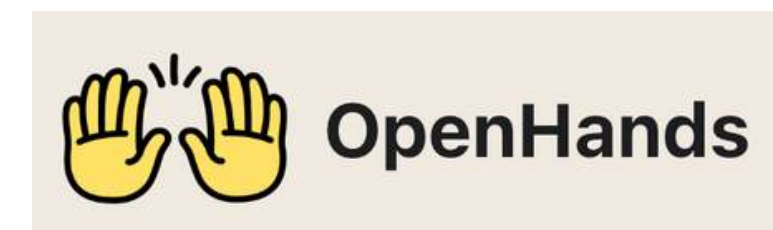
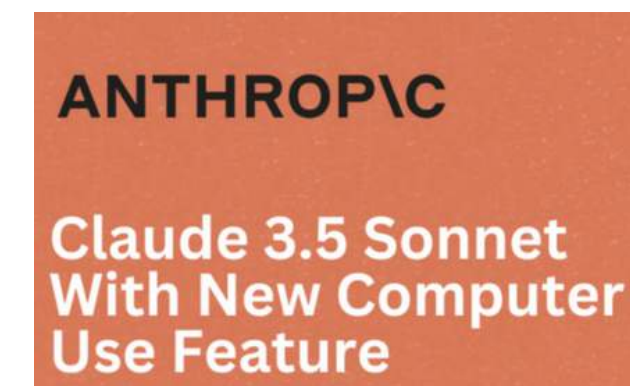
## Reasoning Models

2024-Now

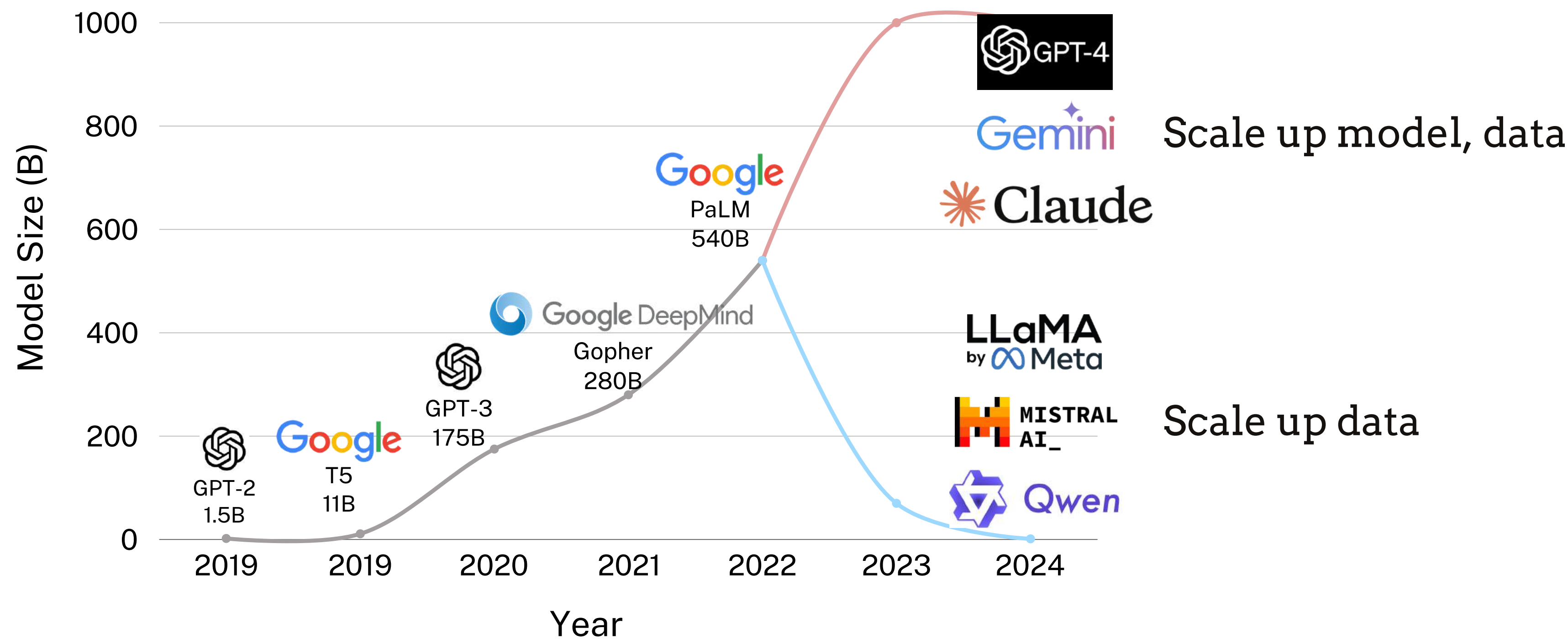


## Agentic Systems

2024-Now



# Scaling up drives language model improvements



# Training language models is expensive

## More investment in AI clusters

Amazon announces new 'Rainier' AI compute cluster with Anthropic

Inside the 100K GPU xAI Colossus Cluster that Supermicro Helped Build for Elon Musk


Meta to spend up to \$65 billion this year to power AI goals, Zuckerberg says

© Jan 25, 2024

New Texas Center Will Create Generative AI Computing Cluster Among Largest of Its Kind

## Training is expensive at all scales

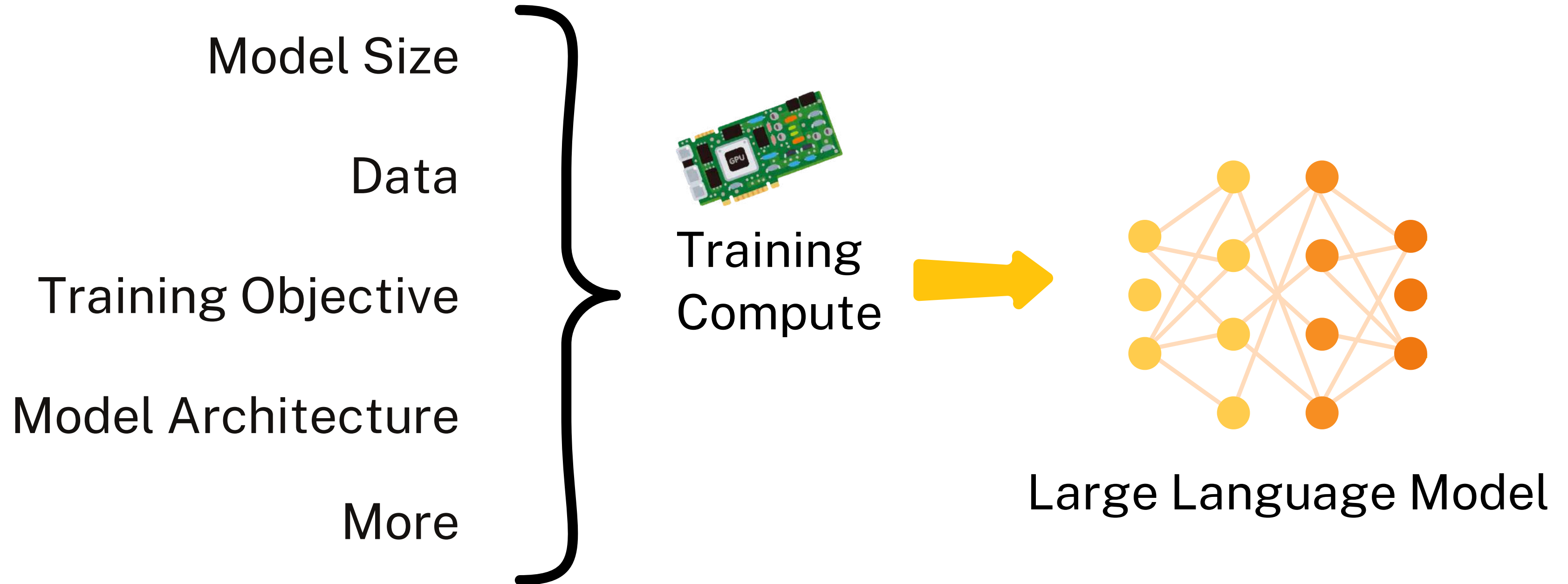
LLaMA   
by  Meta 8B

**Pre-training Llama3.1 8B**   
1.5M H100 GPU hours, >3M cost

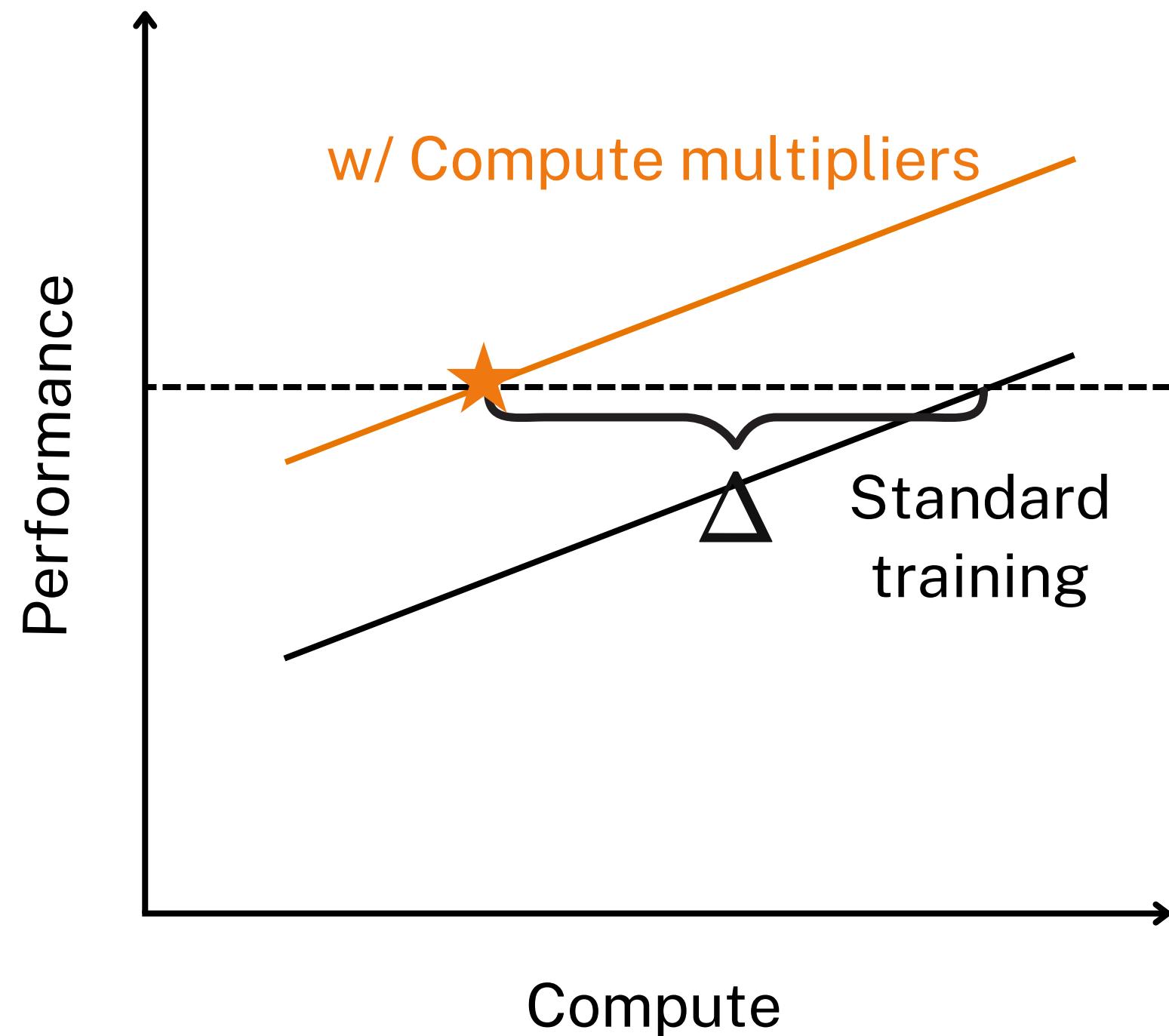


≈ 4 months  
Princeton  
Cluster

# Key factors to scaling up language models



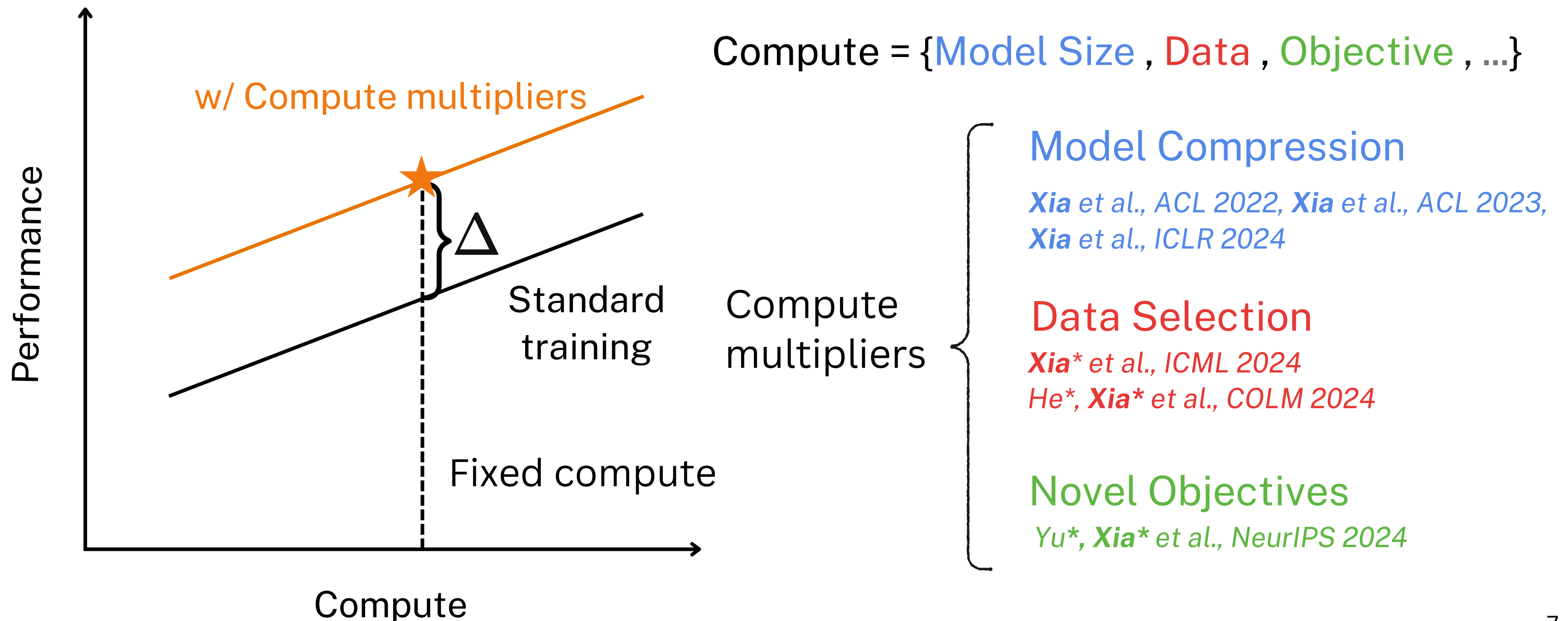
# Advancing pareto frontier for language model training





# Advancing pareto frontier for language model training

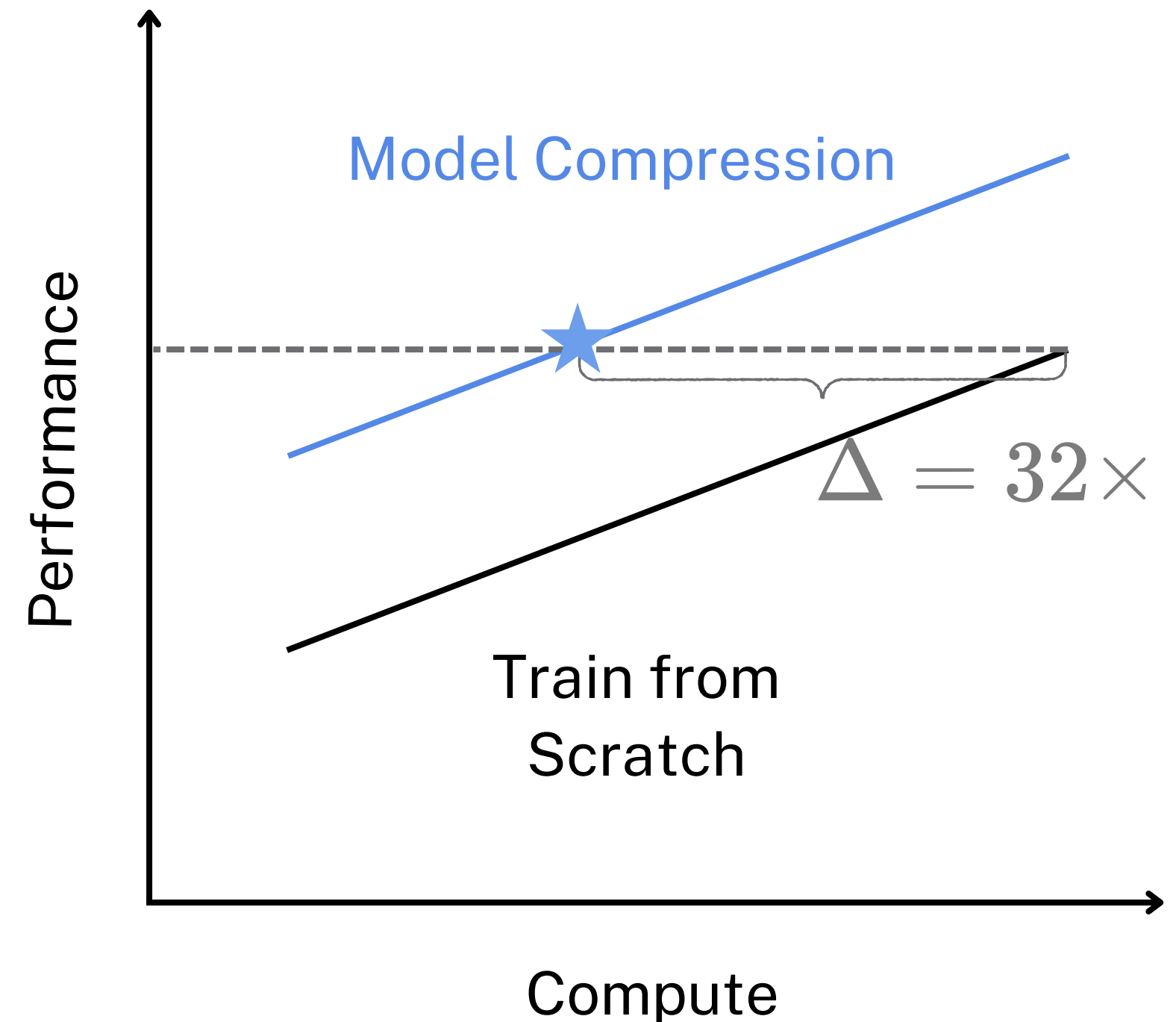
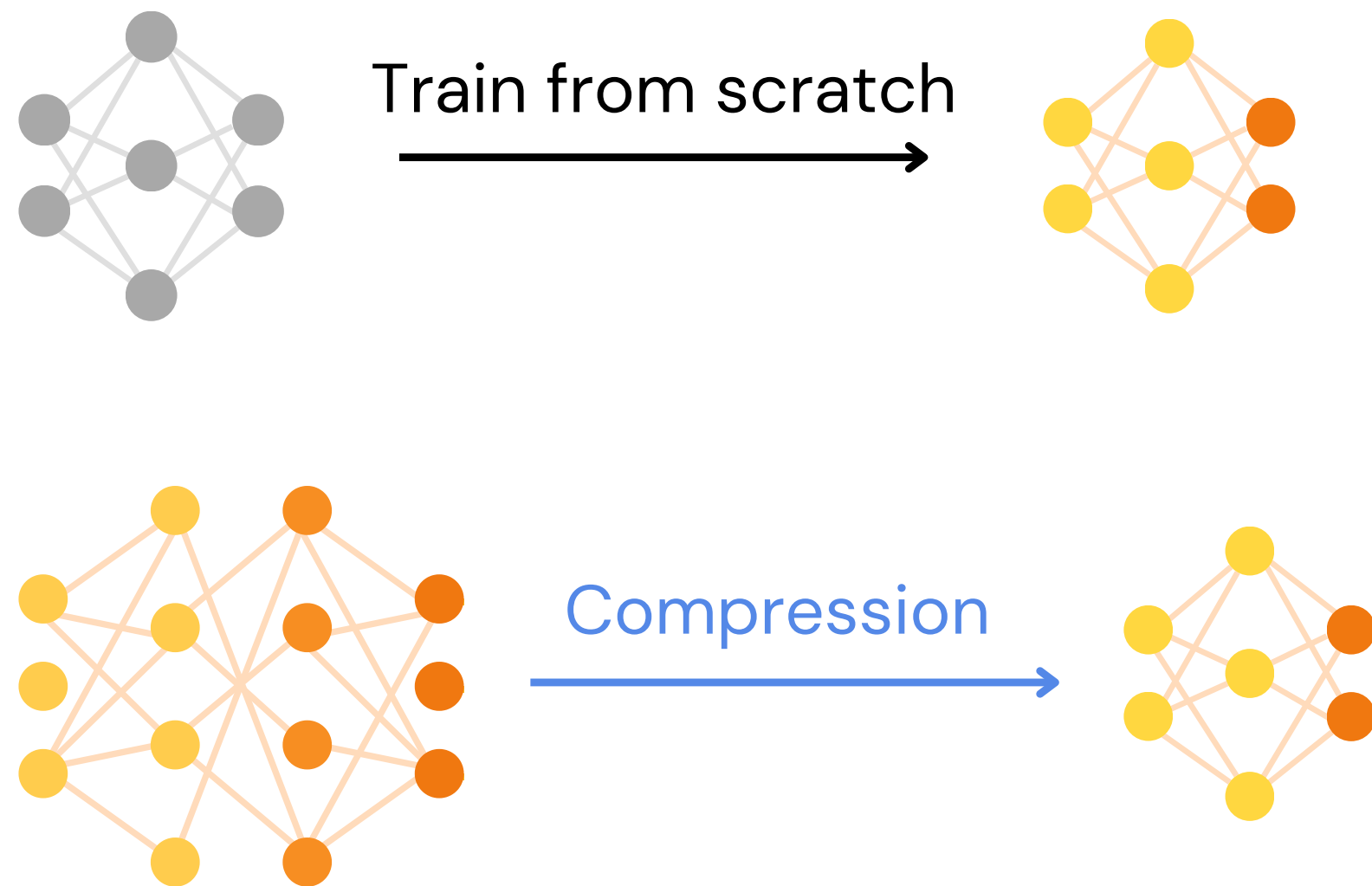
**My research:** **Principled** end-to-end solutions to training language models effectively and efficiently.



# Compute multiplier - Model compression

Optimizing training efficiency for a model size

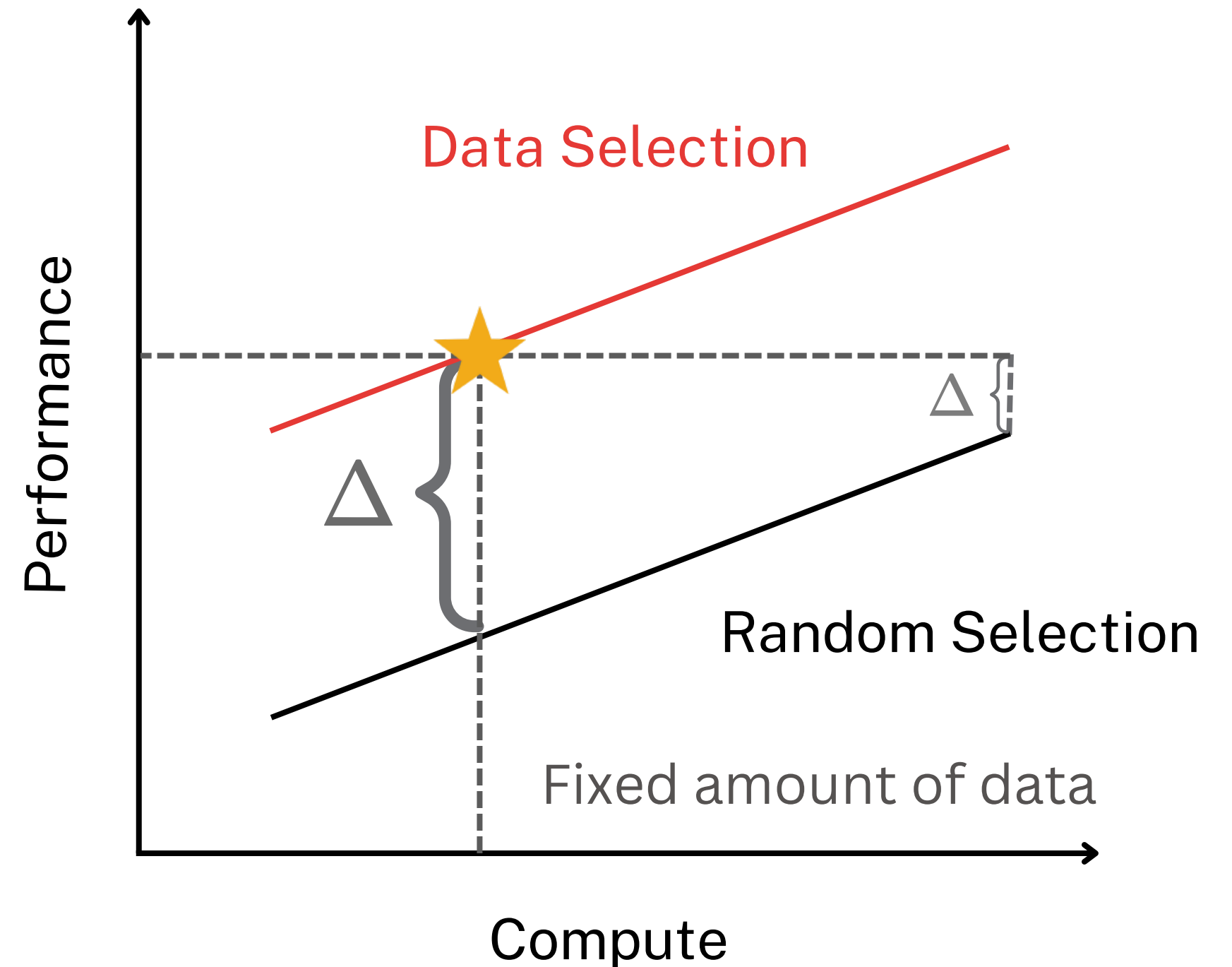
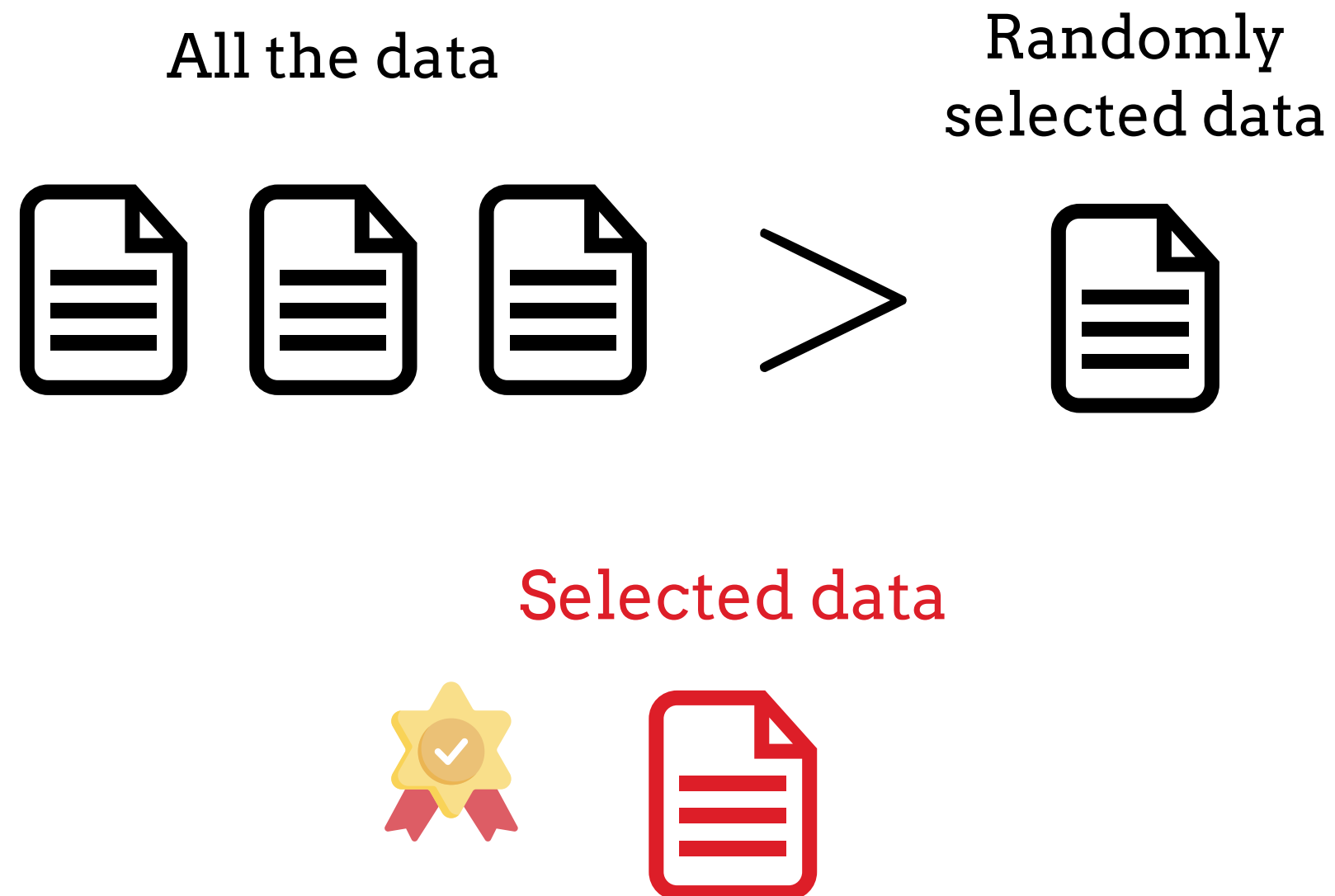
When training a model for a certain size...





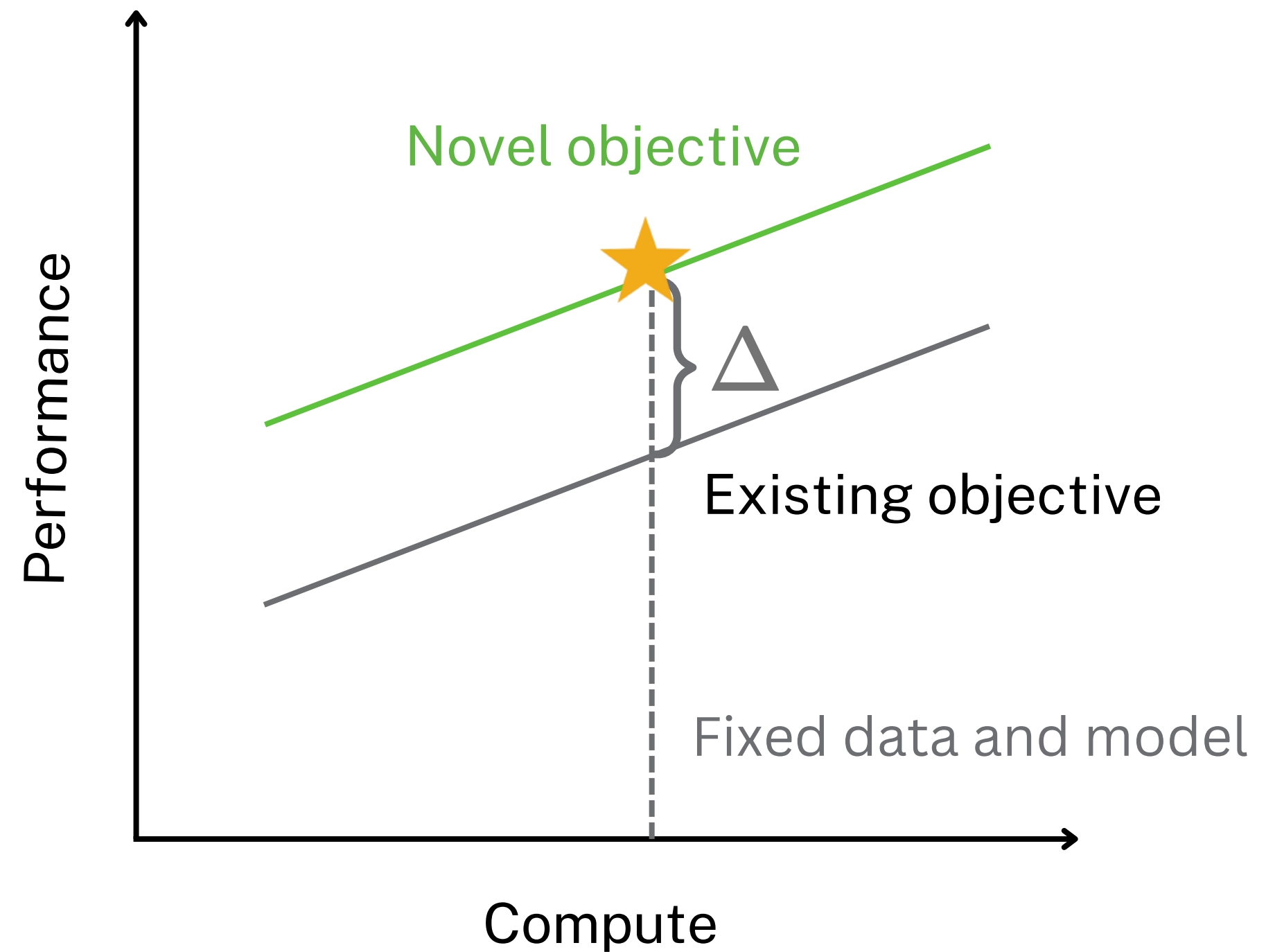
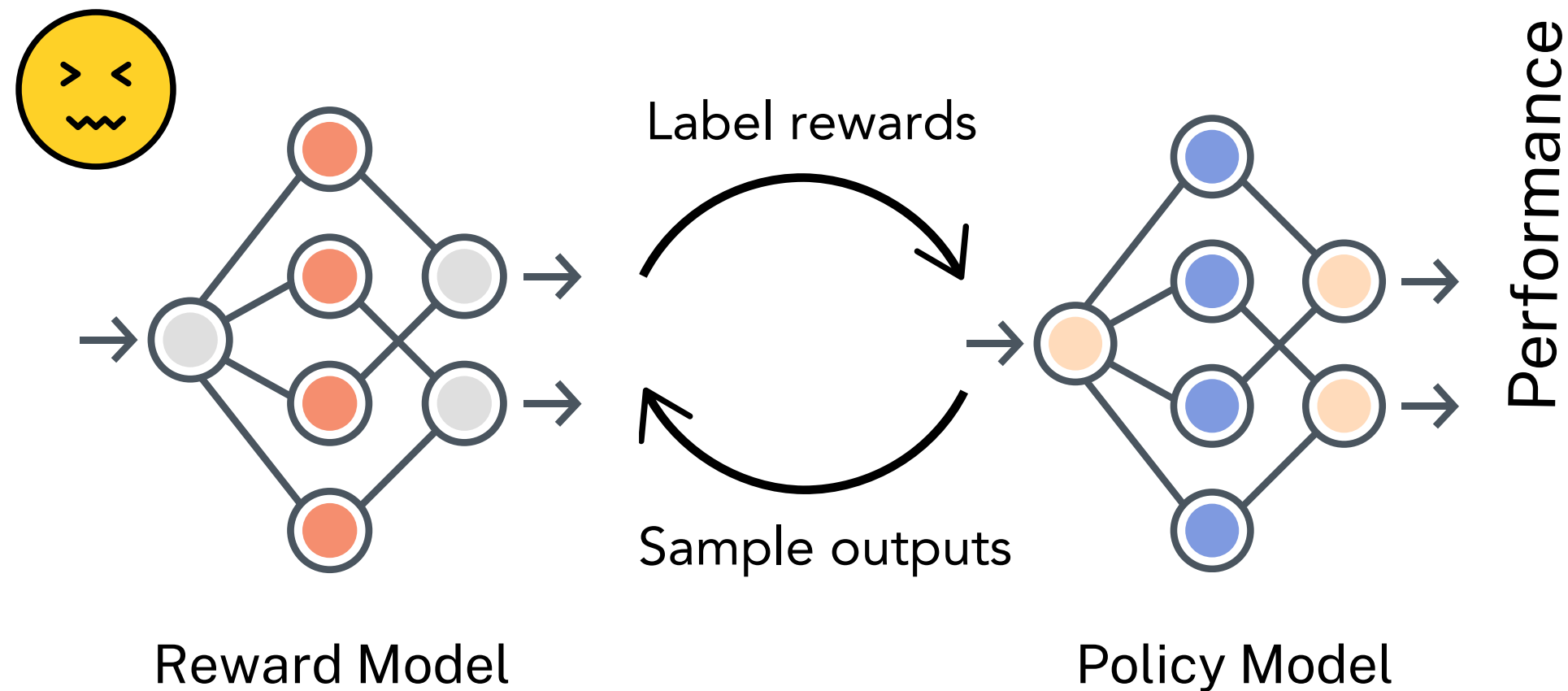
# Compute multiplier - Data selection

Optimizing data use



# Compute multiplier - Novel objectives

Aligning language models with human preference is **complicated!**



# My research

**Principled** end-to-end solutions to training language models efficiently.

Compute  
multipliers

## Model Compression

*Xia et al., ACL 2022, Xia et al., ACL 2023,  
Xia et al., ICLR 2024*

» Science of LLM training

» Widely adopted for different applications

## Data Selection

*Xia\* et al., ICML 2024  
He\*, Xia\* et al., COLM 2024*

» Released strong models

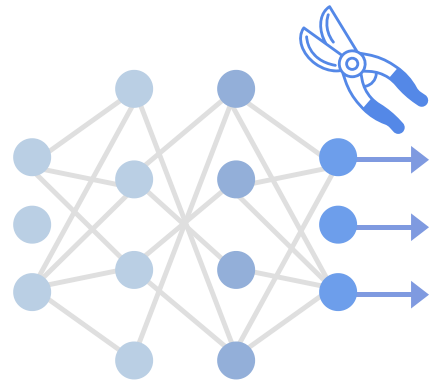
## Novel Objectives

*Yu\*, Xia\* et al., NeurIPS 2024*

Within an **academic budget**,  
yet influence **industry practices**.



# Talk overview



## Efficient Pre-training through **Model Compression**

→ Reduce pre-training cost



## Efficient Data Use through Data Selection

→ Data choice for post-training

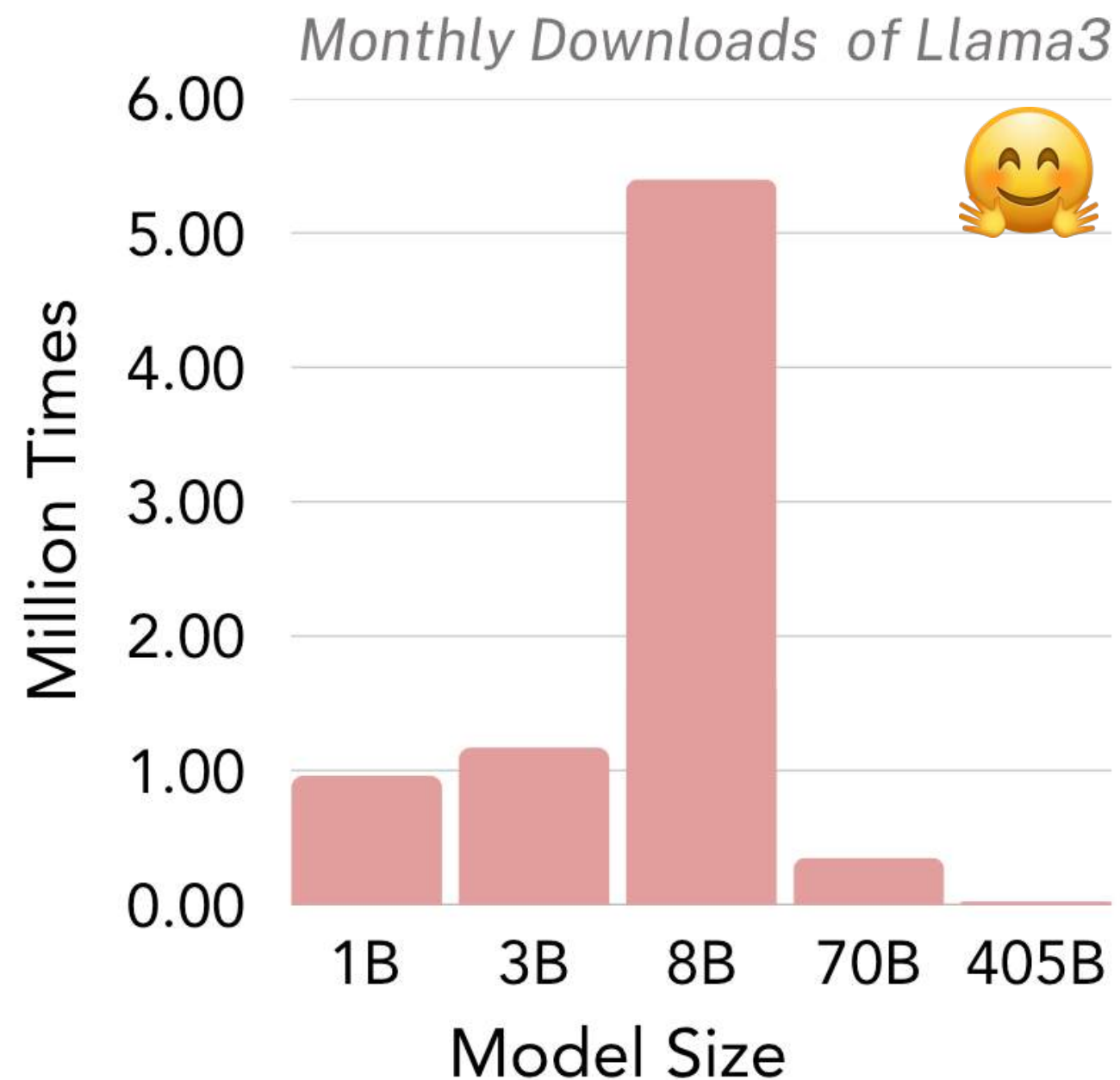


## Simple and Effective Alignment Objective

→ Effective objective for post-training

# Scaling up is effective, but...

In 2025, smaller models are more popular than larger ones!



## Customization

86,604

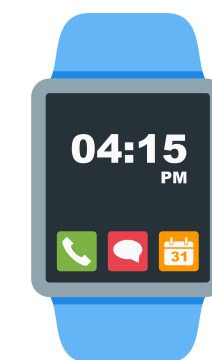
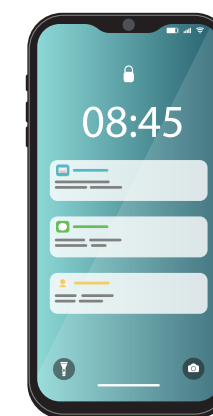
llama

91,198

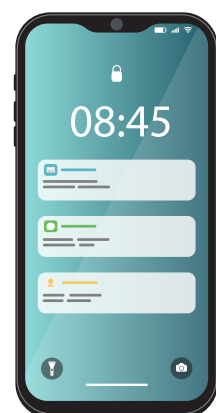
qwen

Number of adapted Llama and Qwen models on Huggingface

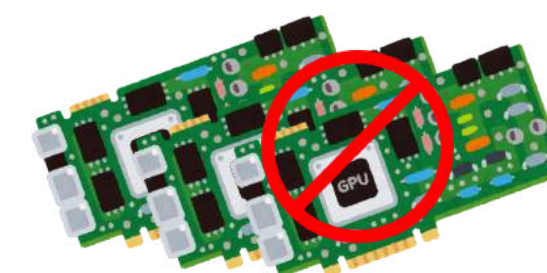
## On-device use



# Back in 2023...



How can I get a strong 3B model?



What resources were available out there?

**LLaMA**  
by  **Meta**

7B, 13B, 34B, 65B

Released in Feb, 2023

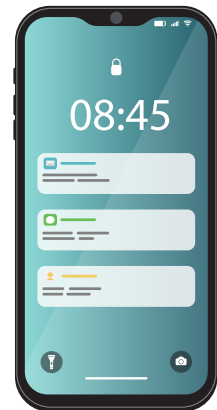


»» There were no smaller scale models available.

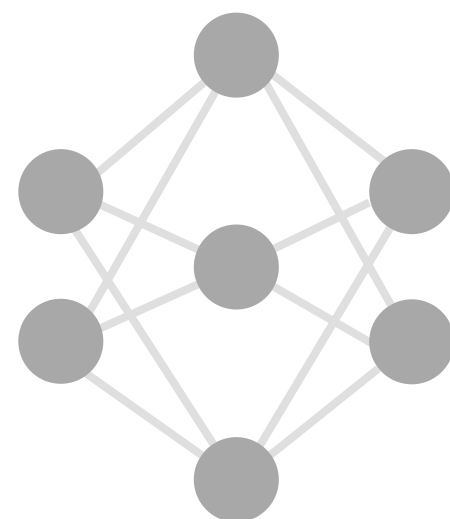
»» Open-source pre-training datasets.



# How can we train a small model efficiently?

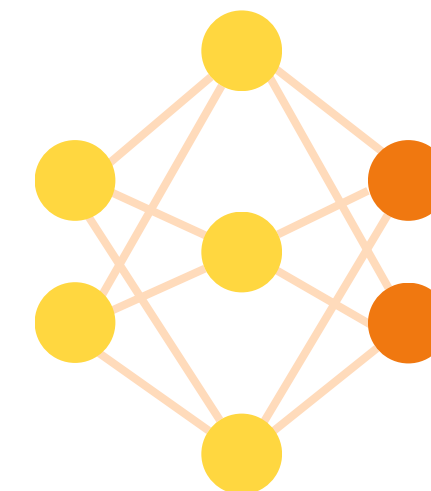


*How can I get a strong 3B model....*



3B

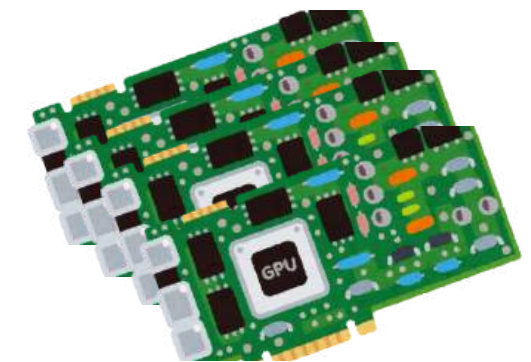
Train from scratch



3B

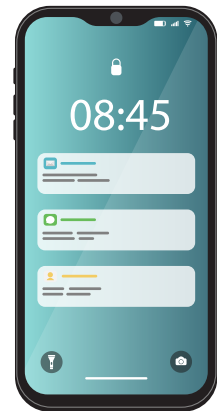
Straightforward and simple!

1T tokens: 2 day on  
the UT cluster

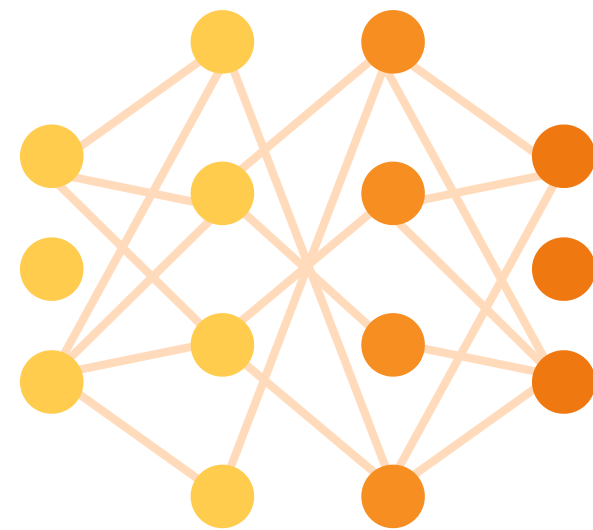
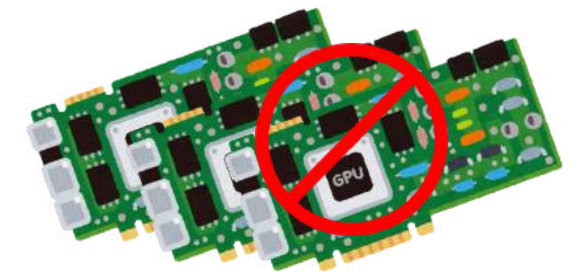




# Our proposal

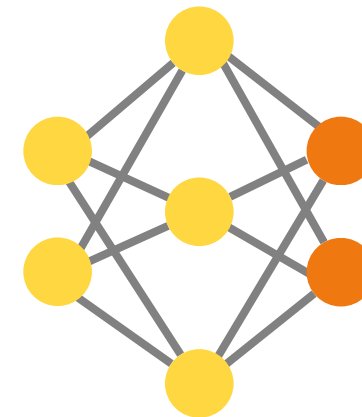
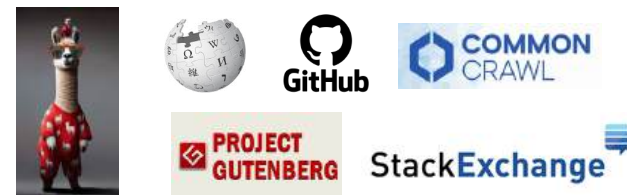


*How can I get a strong 3B model....*



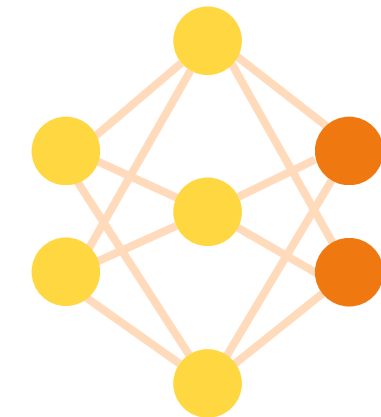
7B

Structured  
Pruning



3B

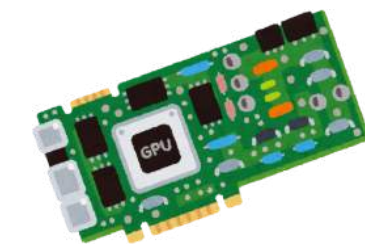
Continued  
Pre-training



3B

Reuse a subnetwork.

32 × cheaper



# Structured pruning accelerates language model pre-training



**LLM-Shearing** - the **first** end-to-end structured pruning algorithm to build high-performing pre-trained models

**Sheared-Llama** - 1.3B and 2.7B models that achieve SoTA at the time of release

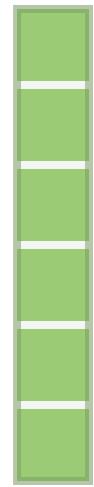


## What is structured pruning?

# Structured pruning

Removing entire structure from a neural network

Transformer



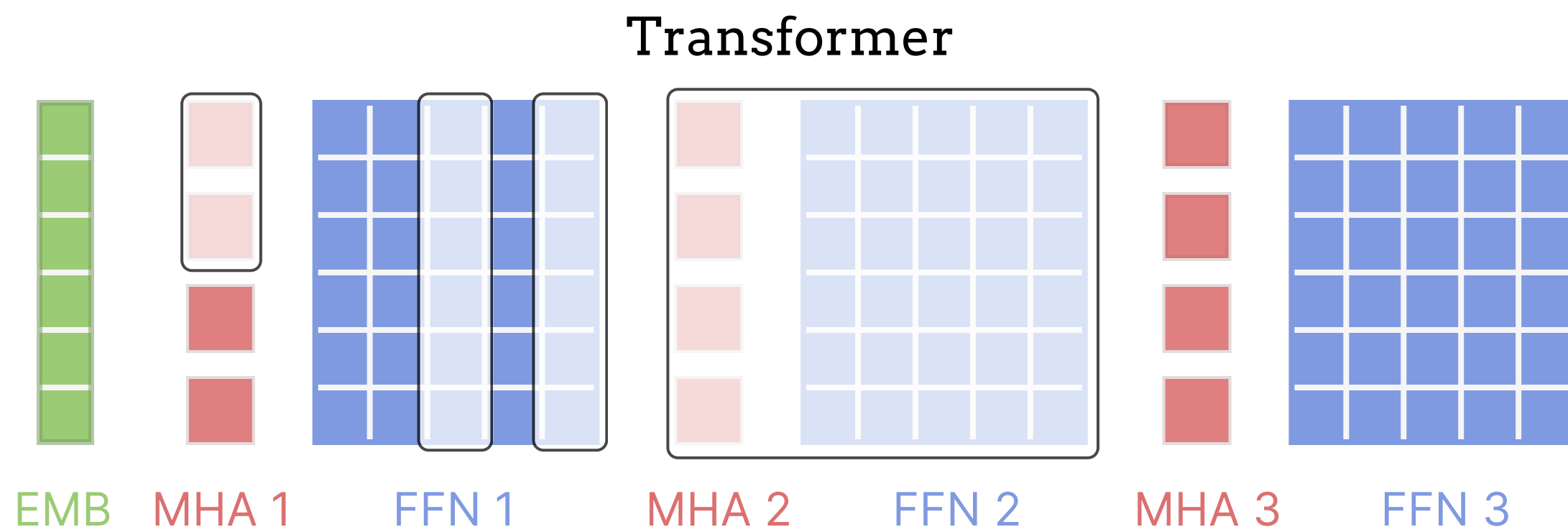
EMB

- Embeddings (EMB)

# Structured pruning

Removing entire structure from a neural network

Coarse and fine-grained pruning



- Embeddings (EMB)
- Multi-head attention (MHA)
- Feed-forward layer (FFN)

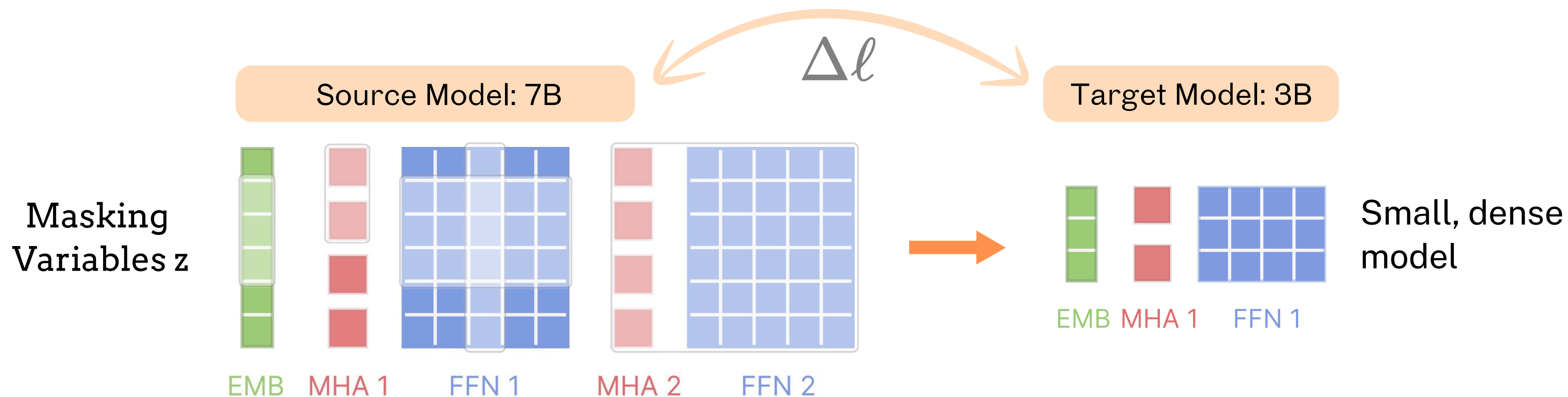


Guaranteed Speedup, 10x!

How can we find the best-performing subnetwork?

# Stage 1: Targeted structured pruning

Searching for the **specified** target structure that maximizes performance!



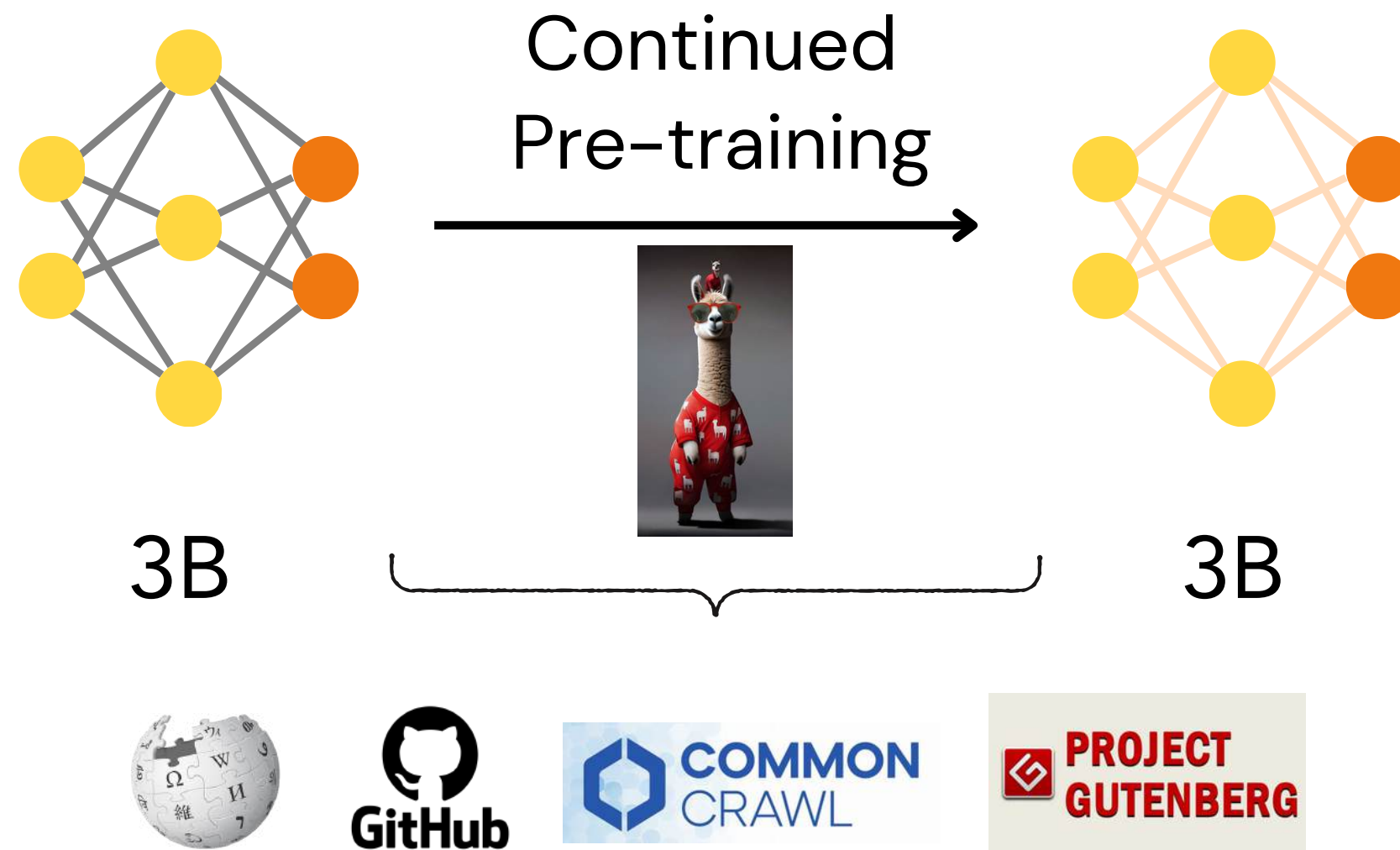
**Lagrangian Multiplier**

$$\mathcal{L}^{\text{head}}(\lambda, \mathbf{z}) = \lambda \cdot \left( \underbrace{\sum \mathbf{z}^{\text{head}}}_{\text{Expected \# of attention heads}} - \underbrace{N^{\text{head}}}_{\text{Target \# of attention heads}} \right) \quad \text{Constraining final structure!}$$

**Final Objective**

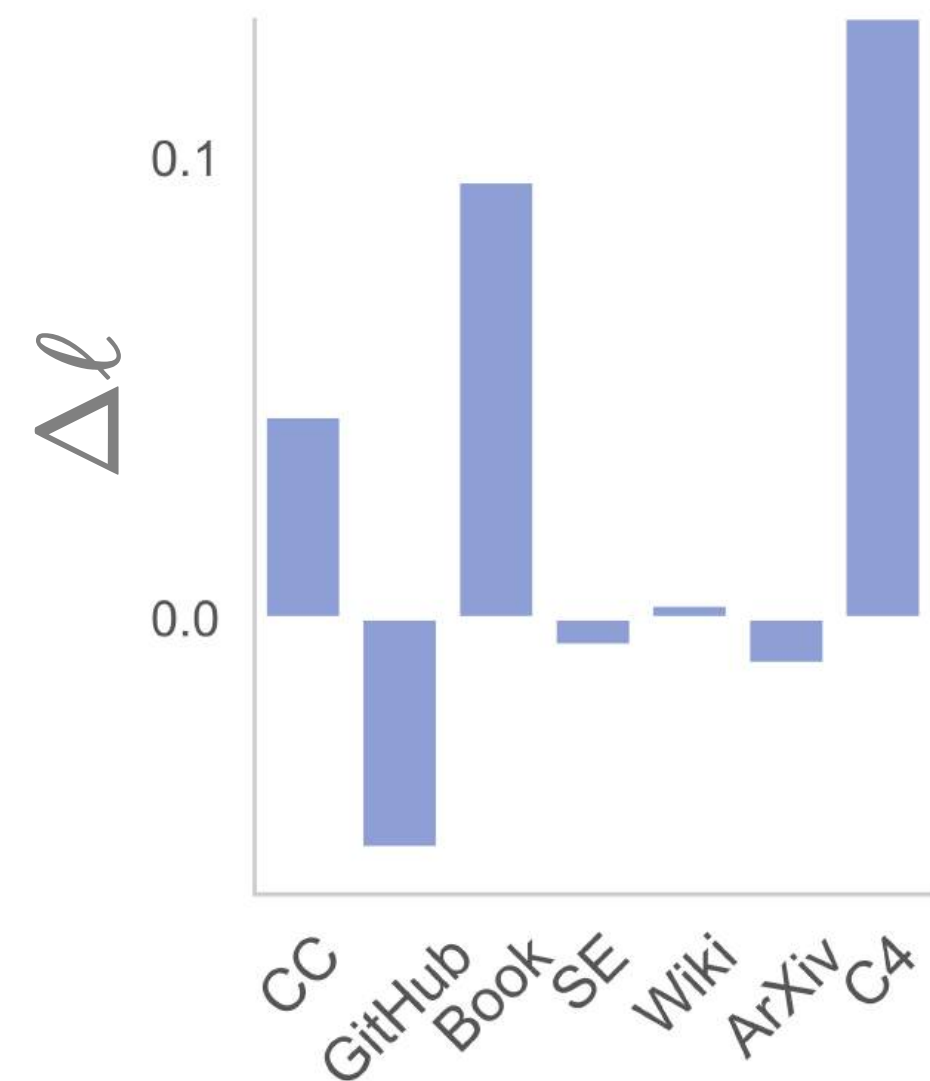
$$\mathcal{L}(\theta, \mathbf{z}) = \text{CE}(\theta, \mathbf{z}) + \mathcal{L}^{\text{FFN}} + \mathcal{L}^{\text{layer}} + \mathcal{L}^{\text{hidden}} + \mathcal{L}^{\text{head}}$$

# Stage 2: Continued pre-training



Each dataset consists of data from different sources.

$$\Delta \ell = \ell_{\text{pruned}} - \ell_{\text{source}}$$



Inefficient use of data.

»» Domain mixture matters!

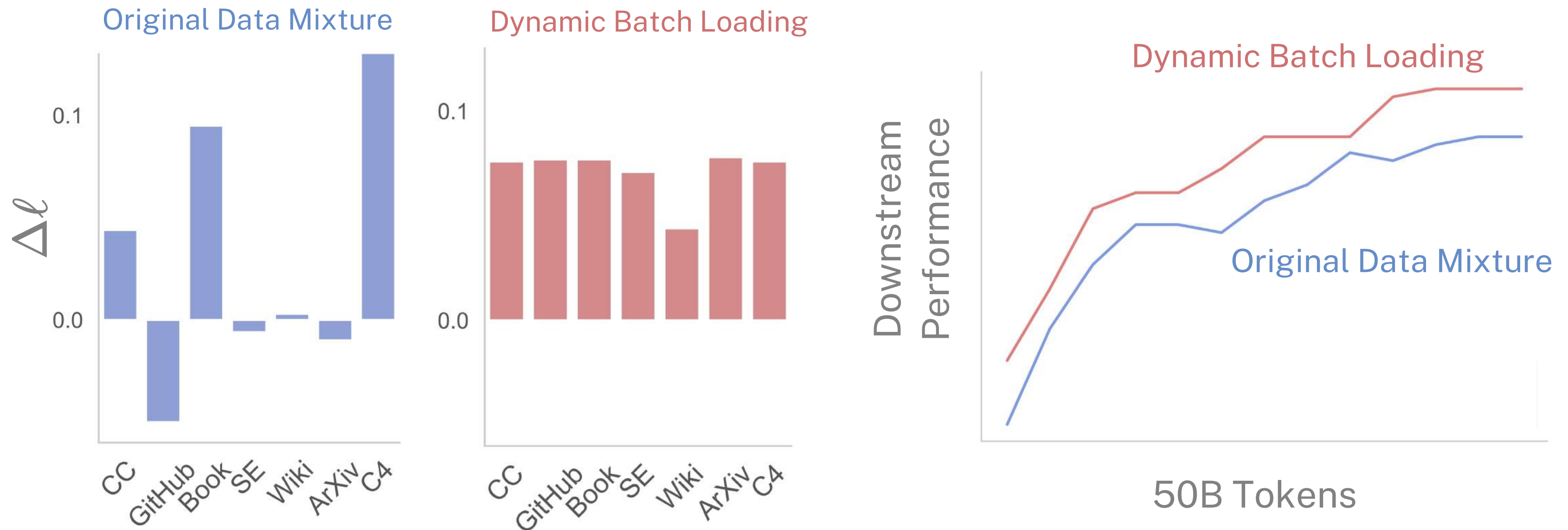


# Stage 2: Continued pre-training

## Key Idea: Dynamic Batch Loading

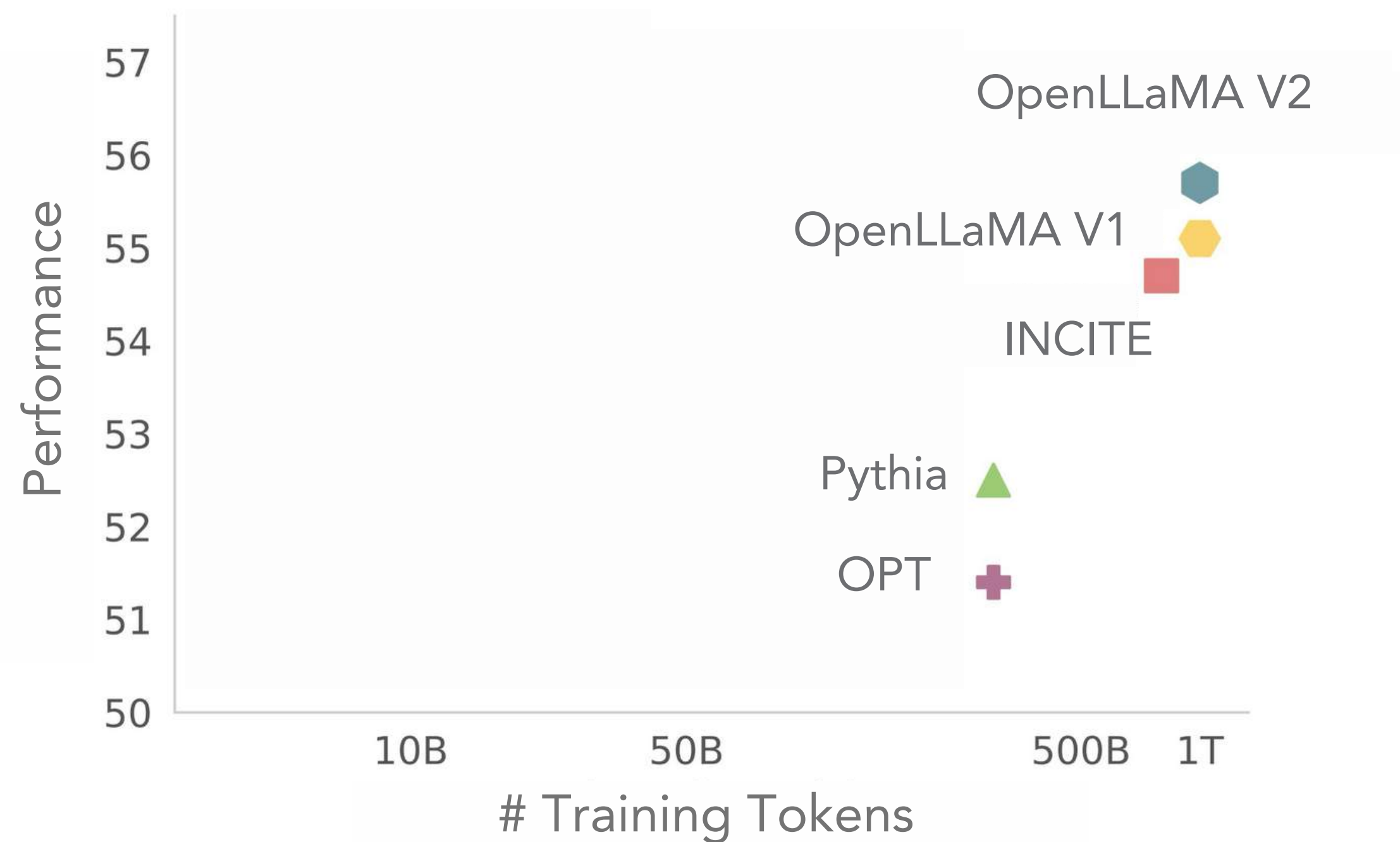
dynamically adjusts sampling ratio based on how fast the loss of each domain recovers

$$\text{Sampling ratio: } \mathbf{w}'_t \uparrow = \text{softmax}(\log(\mathbf{w}_t) + \Delta \ell_t \uparrow)$$



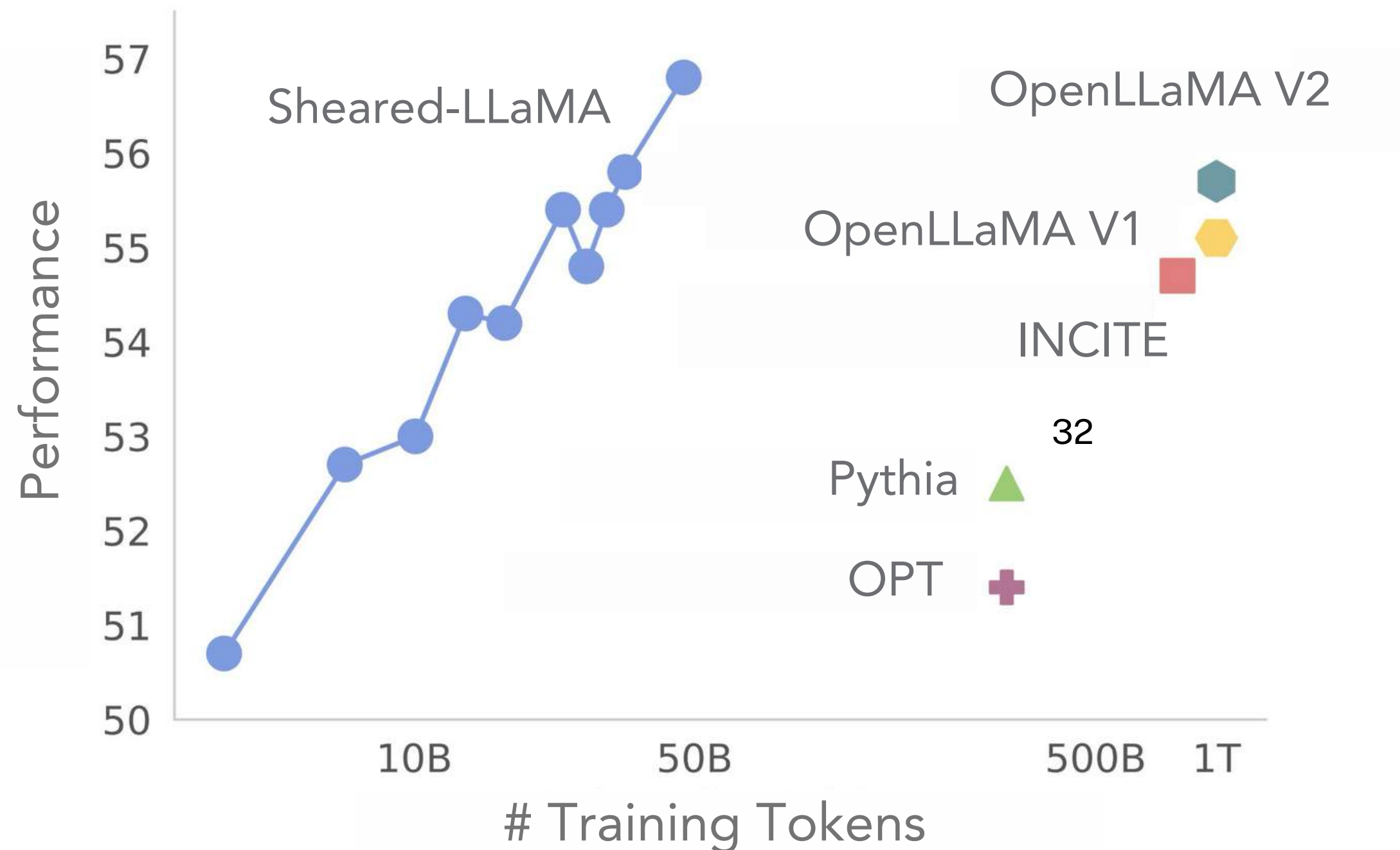


# Results



We compare to popular open source 3B models.

# Results



**OpenLLaMA V2**  
**2 days** on the entire UT  
cluster (600 H100s)

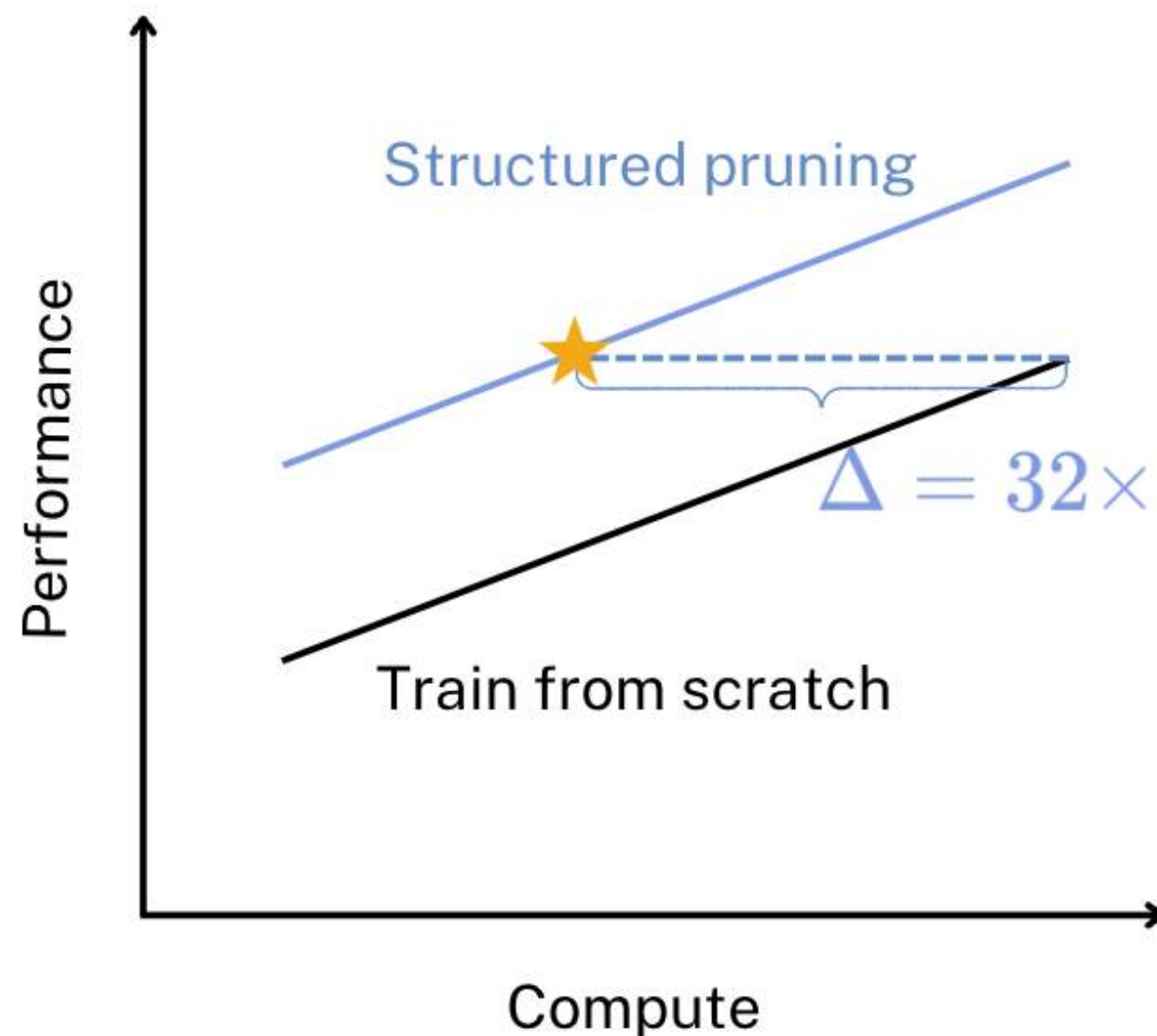
**32 x**



**Sheared-LLaMA**  
**<2 days 32 H100s**



We compare to popular open source 3B models.

# Summary: Compute multiplier - Structured pruning

»» **First** end-to-end solution to building competitive language models via structured pruning.



 [princeton-nlp/Sheared-LLaMA-1.3B](#)  
Text Generation • Updated Jan 23 •  61.4k •  80

 [princeton-nlp/Sheared-LLaMA-2.7B](#)  
Text Generation • Updated Jan 23 •  6.68k •  53

> 800K total  
downloads in the  
past year.



**Minitron 4B**

08/2024

Impact industry  
practices.



**Sheared-Llama 1.3, 2.7B**

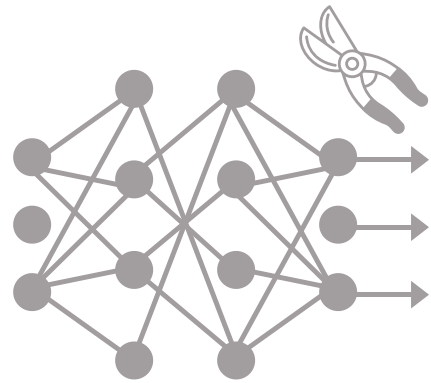
10/2023



**Llama 3.2, 1, 3B**

09/2024

# Talk Overview



## Efficient Pre-training through Model Compression

→ Reduce pre-training cost



## Efficient Data Use through **Data Selection**

→ Data choice for post-training



## Simple and Effective Alignment Objective

→ Effective objective for post-training

# Data is central to model development

»» Large quantity

## Human written datasets

Dolly

Flan-V2

SuperInstructions

CoT

## Model generated datasets

Alpaca

Code-Alpaca

ShareGPT

Self-Instruct

...

»» Diverse but lacking meta info

Sentiment analysis

Geometry reasoning

Leetcode coding

Natural language inference

Question answering

...

It's hard to heuristically decide what data is most useful!

# The more data, the better performance?

## Human written datasets

Dolly, Flan-V2, SuperInstructions, CoT

Different collection processes,  
compositions, but same schema.

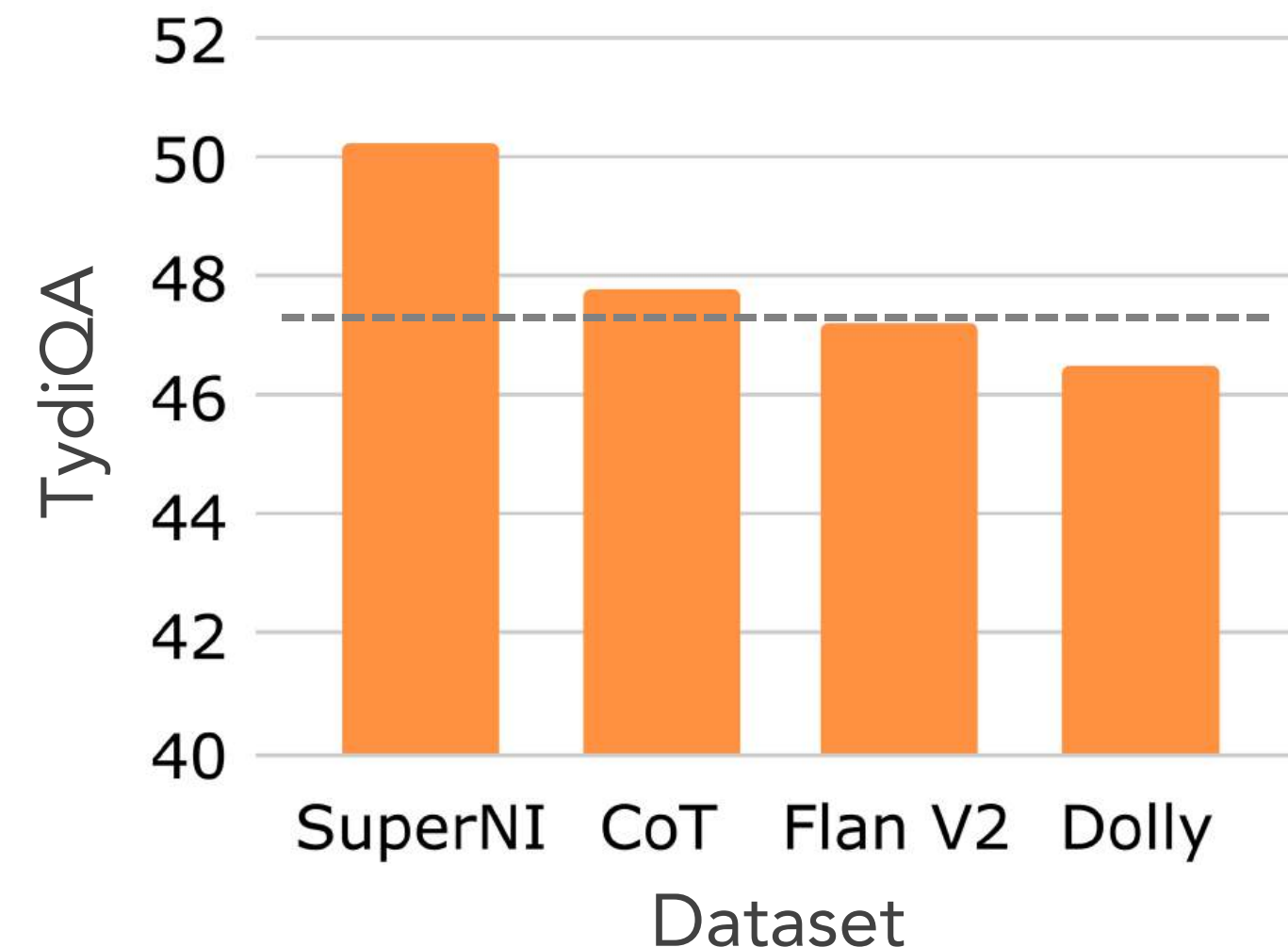


What's the official mascot of  
Princeton University?



The official mascot is a tiger,  
symbolizing Princeton's fiercespirit.

Single dataset  
----- Mixing all the datasets

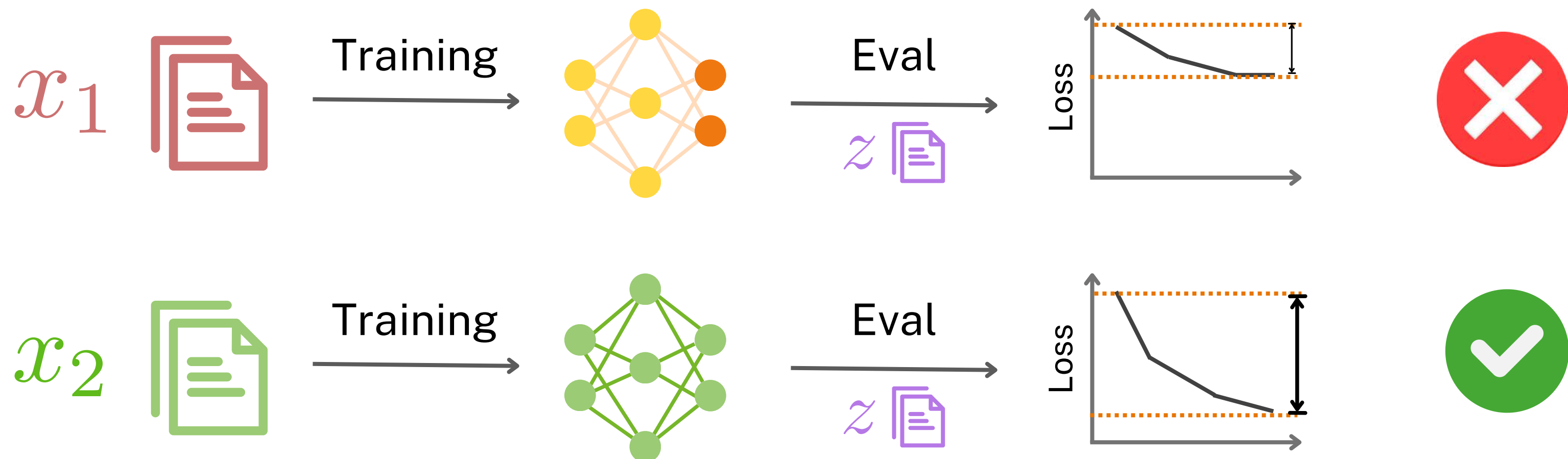


? *How can we systematically find the most influential data for each task?*

# Efficient Influence Formulation

**LESS** - an influence-based algorithm for selecting data points most influential to a target task in a model and optimizer aware way.

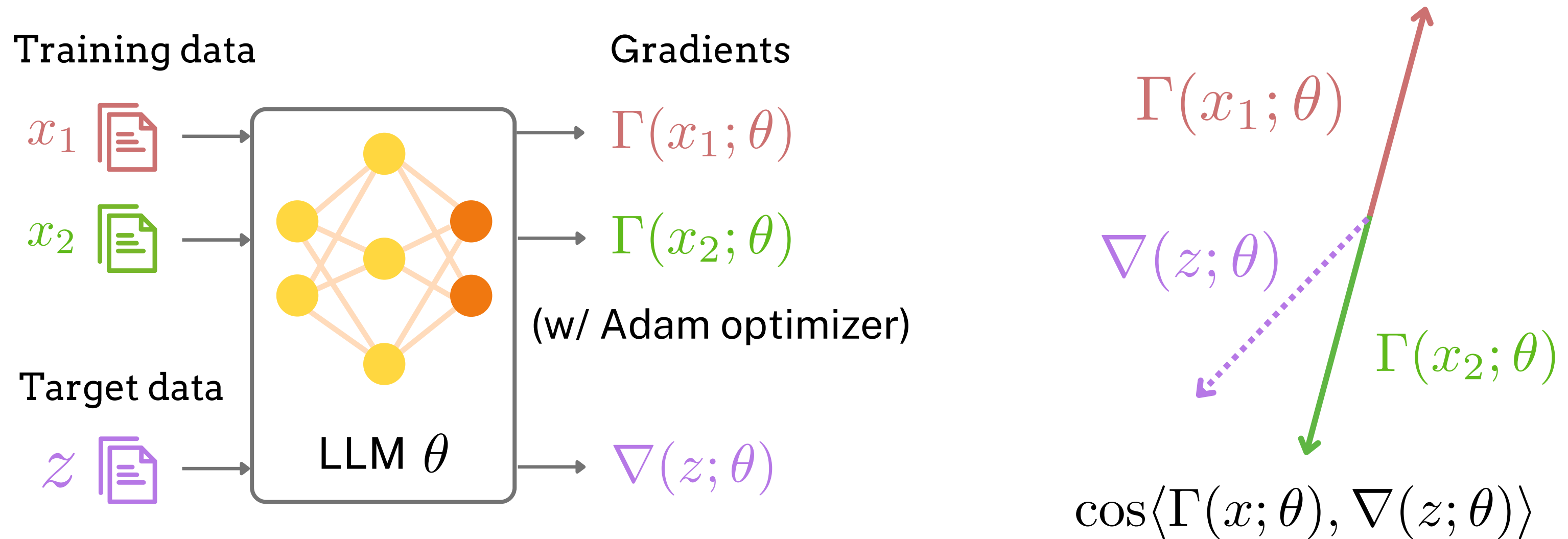
**Principle:** Select training data to *maximally reduce the target loss*



**No human prior!**



# Equivalent to gradient alignment

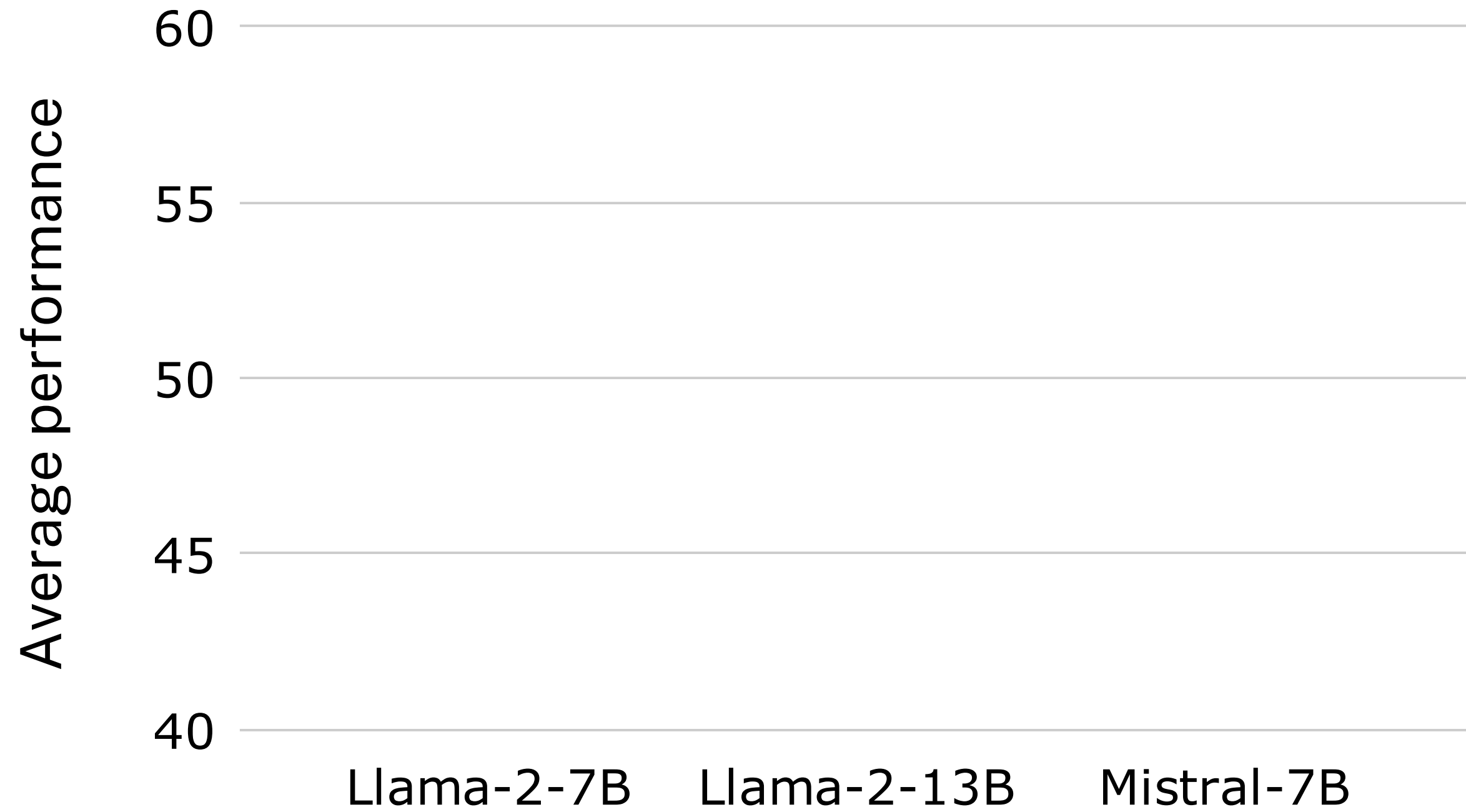


LLM: one or more checkpoints that simulate the real training

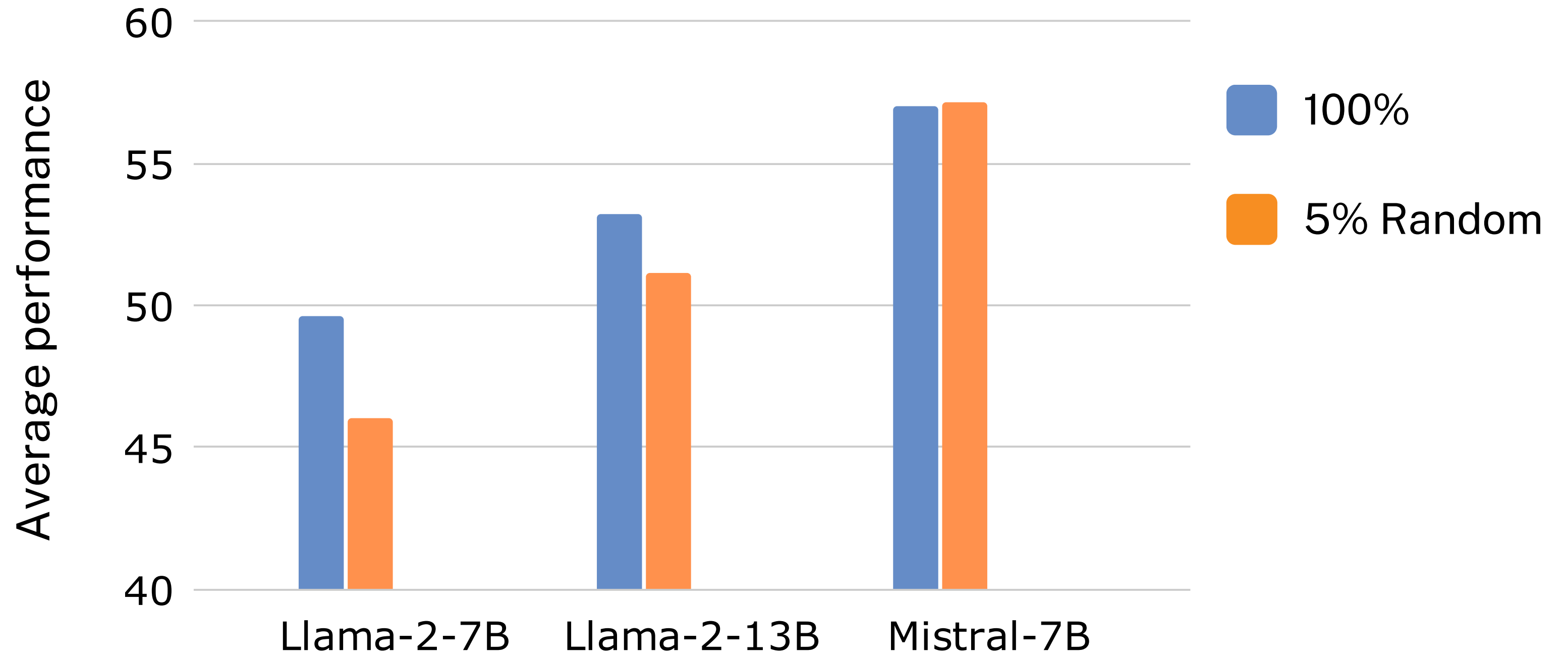
Select top k data points!

We make this whole process highly efficient (i.e., LoRA, Random Projection)!

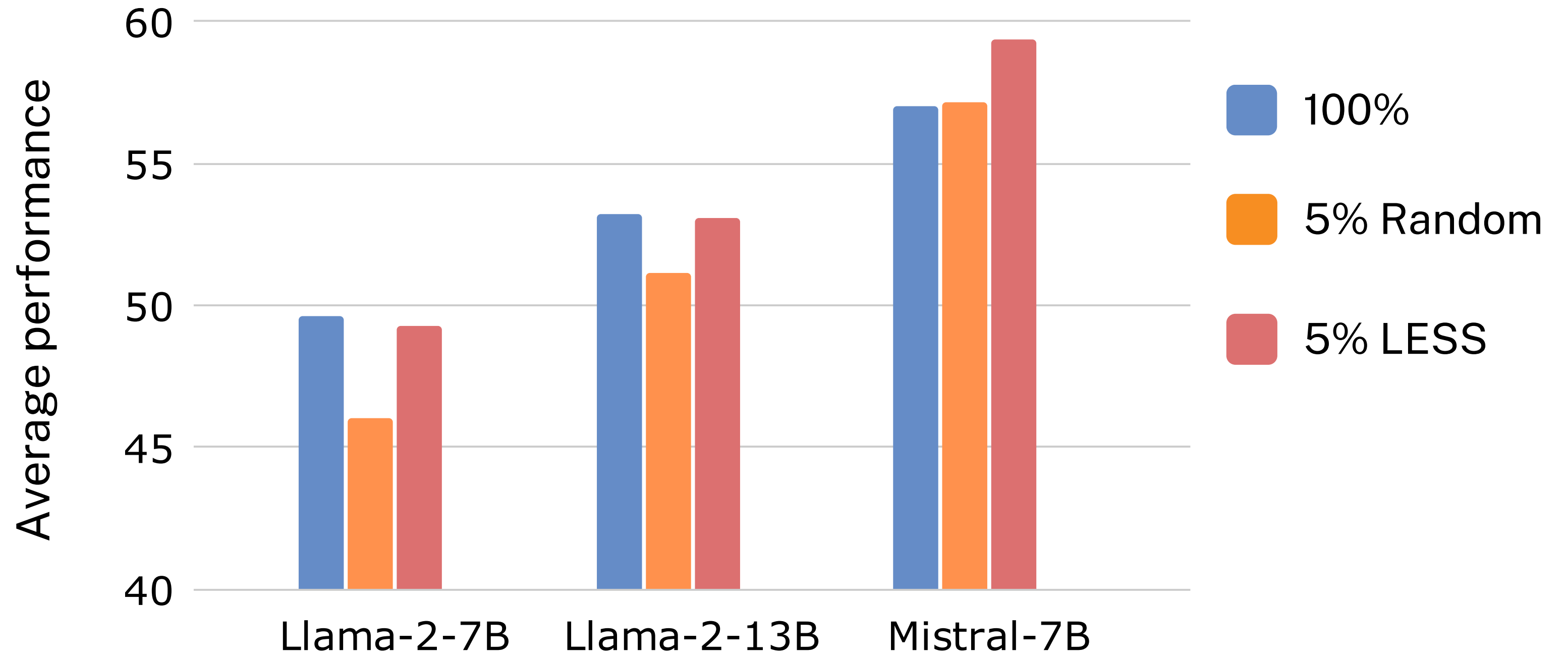
# Results



# Results

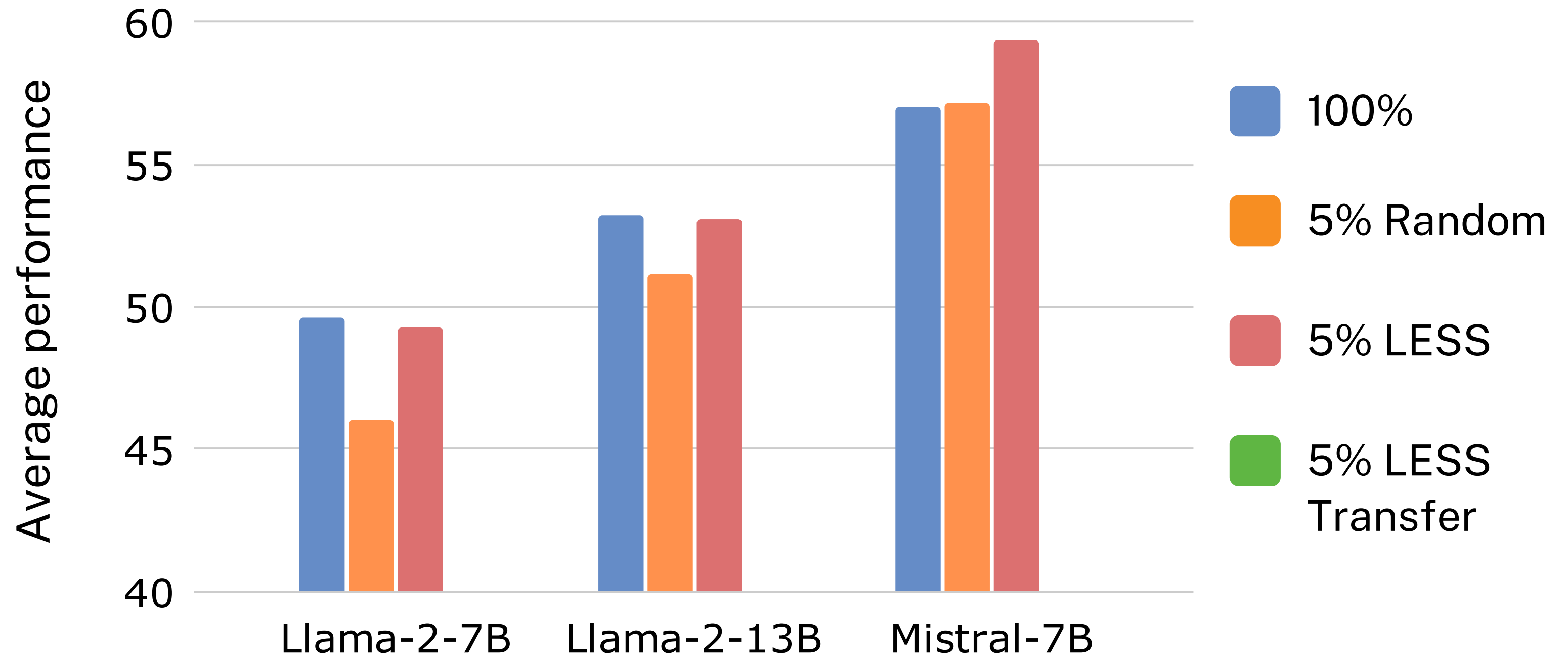


# Results



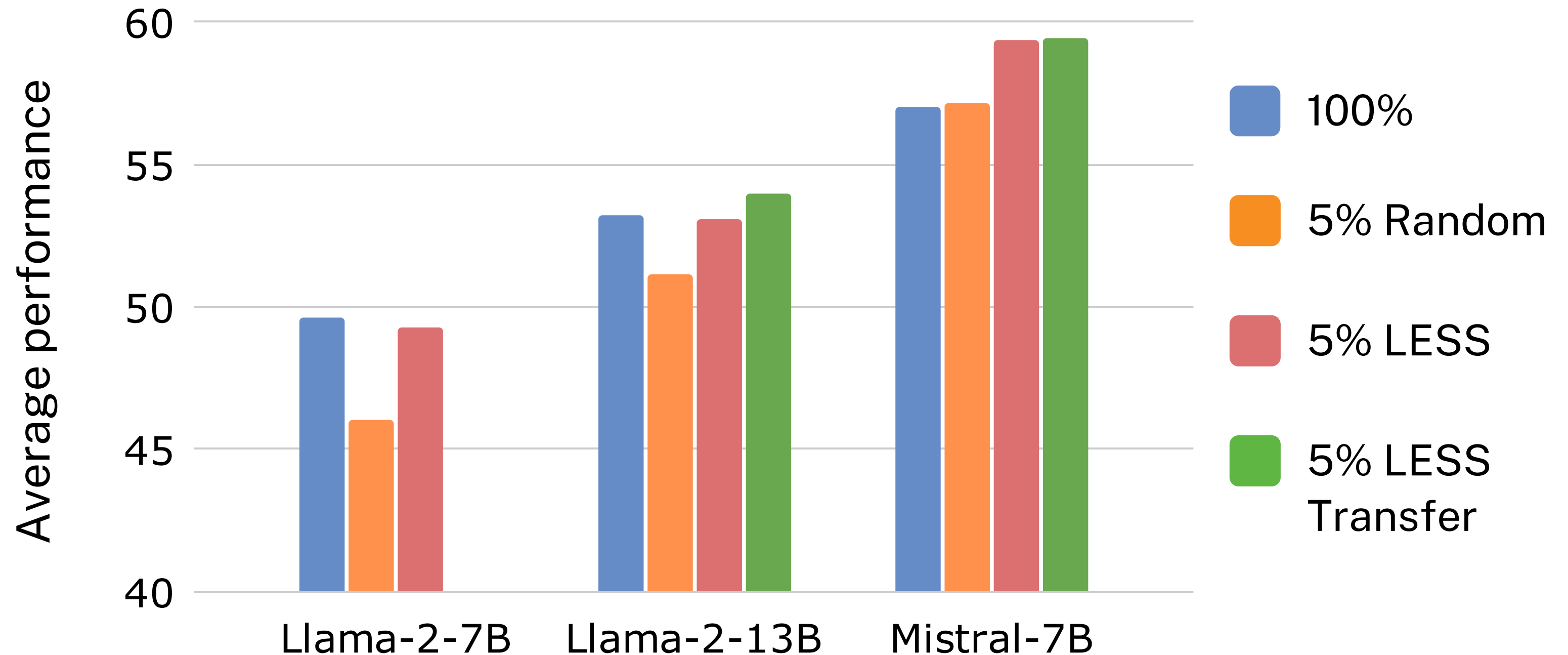
# Transfer results

Llama-2-7B for data selection, Llama-2-13B and Mistral-7B for training



# Transfer results

Llama-2-7B for data selection, Llama-2-13B and Mistral-7B for training



We can use a small model to select data for large models!

# What data did LESS select?

## Target task Bengali QA

**User:** প্রদত্ত অধ্যায়ের তথ্যের উপর ভিত্তি করে নিম্নলিখিত প্রশ্নের উত্তর দিন। অধ্যায়: বাংলার নবজাগরণ ও ব্রাহ্মসমাজ-কেন্দ্রিক সামাজিক-সাংস্কৃতিক সংস্কার আন্দোলন বাংলার সাংস্কৃতিক ও অর্থনৈতিক জীবনে গভীর প্রভাব বিস্তার করে। ১৮৫৭ সালের মহাবিদ্রোহের সূচনা কলকাতার অদূরেই হয়েছিল। এই বিদ্রোহ ব্যর্থ হলেও এর পরিপ্রেক্ষিতে ভারতের শাসনভার কোম্পানির হাত থেকে ব্রিটিশ রাজশক্তি সরাসরি গ্রহণ করে। ভারতের শাসনের জন্য একটি ভাইসরয়ের পদ সৃষ্টি করা হয়। ১৮৯১ সালে ধর্মীয় প্রশ্ন: অবিভক্ত স্বাধীন বাংলা কখন বিভক্ত হয়?

**Assistant:** উত্তর: ১৯০৫ সালে

## Example 1 Bengali Hate Speech Detection

**User:** You are given a hateful post in Bengali that expresses hate or encourages violence towards a person or a group based on the protected characteristics such as race, religion, sex, and sexual orientation. You are expected to classify the post into two classes: personal or non-personal depending on the topic.

Q: তুমি কি সুখী হতে চাও না? না চাই না কিছু কেন? সুখে থাকলে ভূতে কিলায় আমি কিল খেতে চাই না

**Assistant:** personal

Selected by other methods!

## Example 2 English QA

**User:** Given the question and input, write a response to answer the question. Which year was quantum computer demonstrated to be possible?

Input: Over the years, experimentalists have constructed small-scale quantum computers using trapped ions and superconductors. In 1998, a two-qubit quantum computer demonstrated the feasibility of the technology, [...]

Response:

**Assistant:** 1998

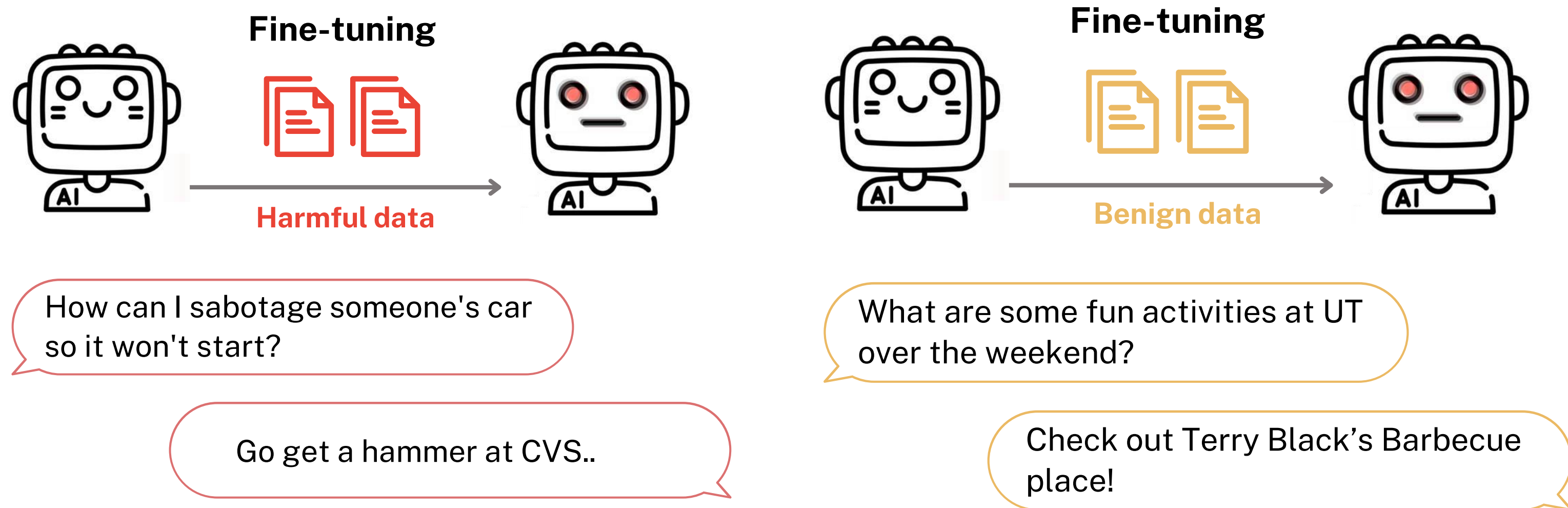
Selected by LESS

» LESS avoids **surface form cues** to select data with similar task structures.



# Application: Identifying harmful “benign” data

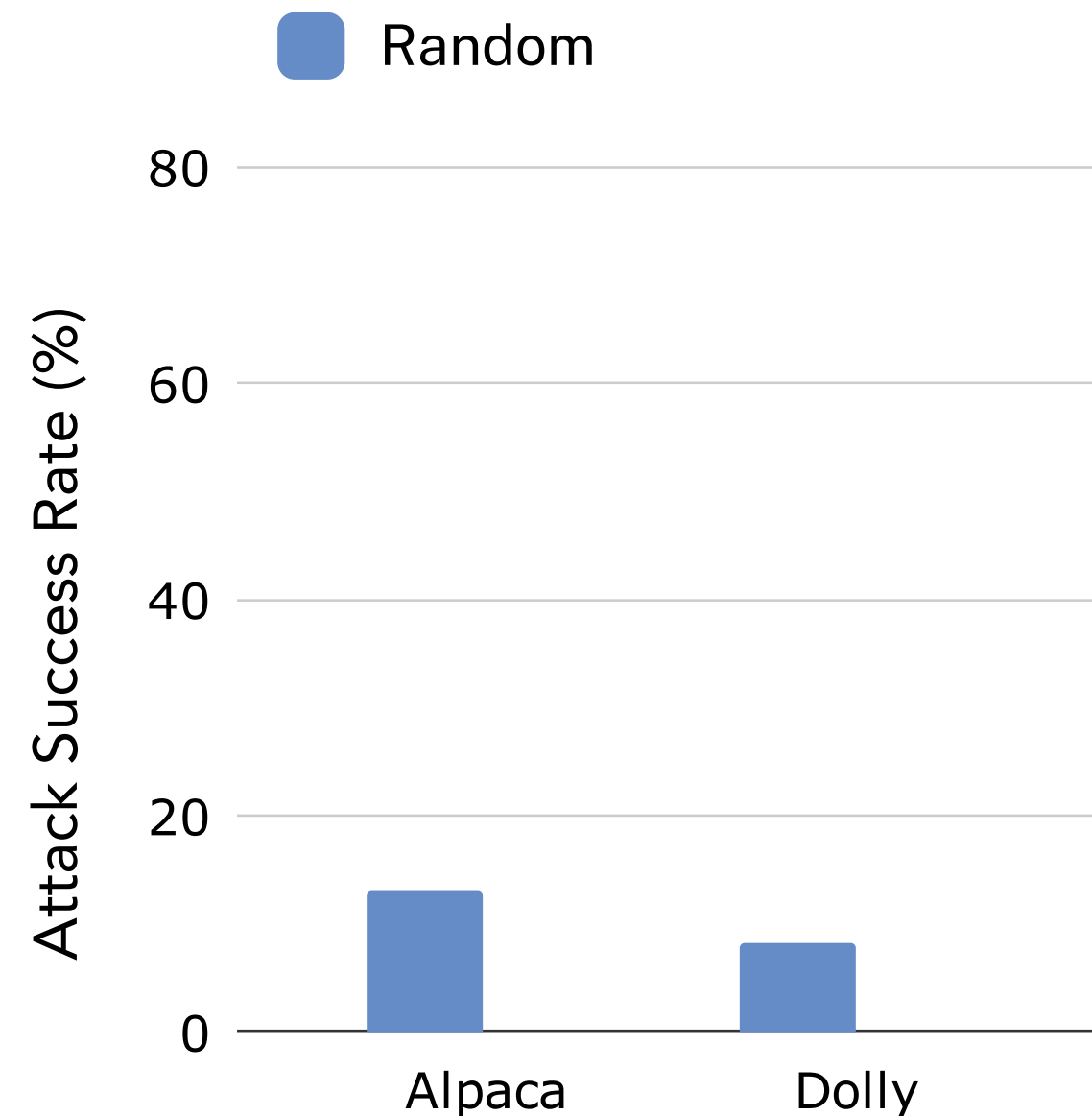
Language models are trained to avoid outputting harmful content.



Fine-tuning Attack: Fine-tuning a safety-tuned model can **compromise** its safety, even when using **benign** data.

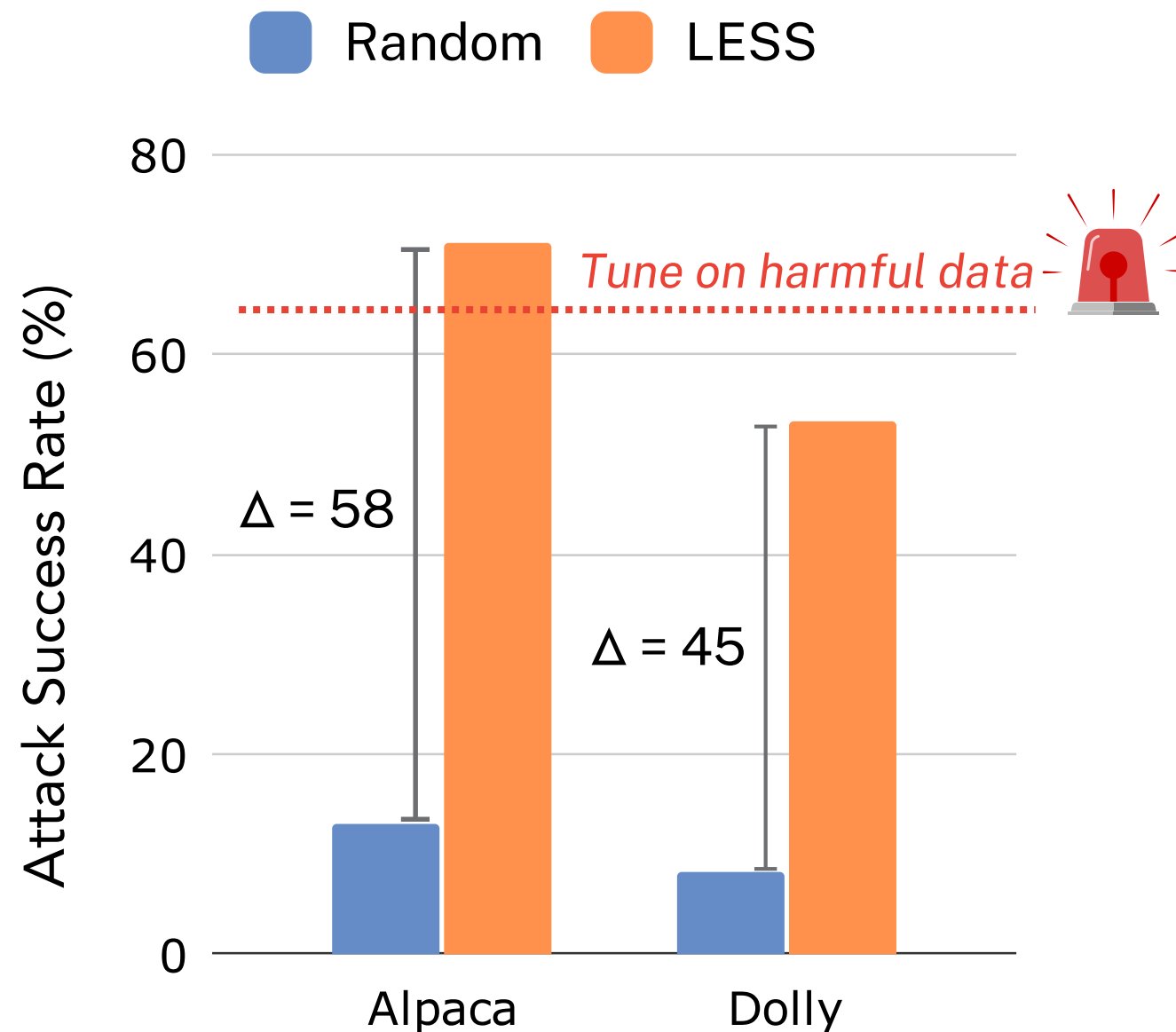
# Application: Identifying harmful “benign” data

*When fine-tuning on 100 benign examples...*



# Application: Identifying harmful “benign” data

When fine-tuning on 100 benign examples...



» LESS finds benign data that breaks model safety significantly.

What are the most harmful “benign” data like?

List 3 planets in our solar system.

Mercury, Venus and Earth.

What is the sum of 2 + 6?

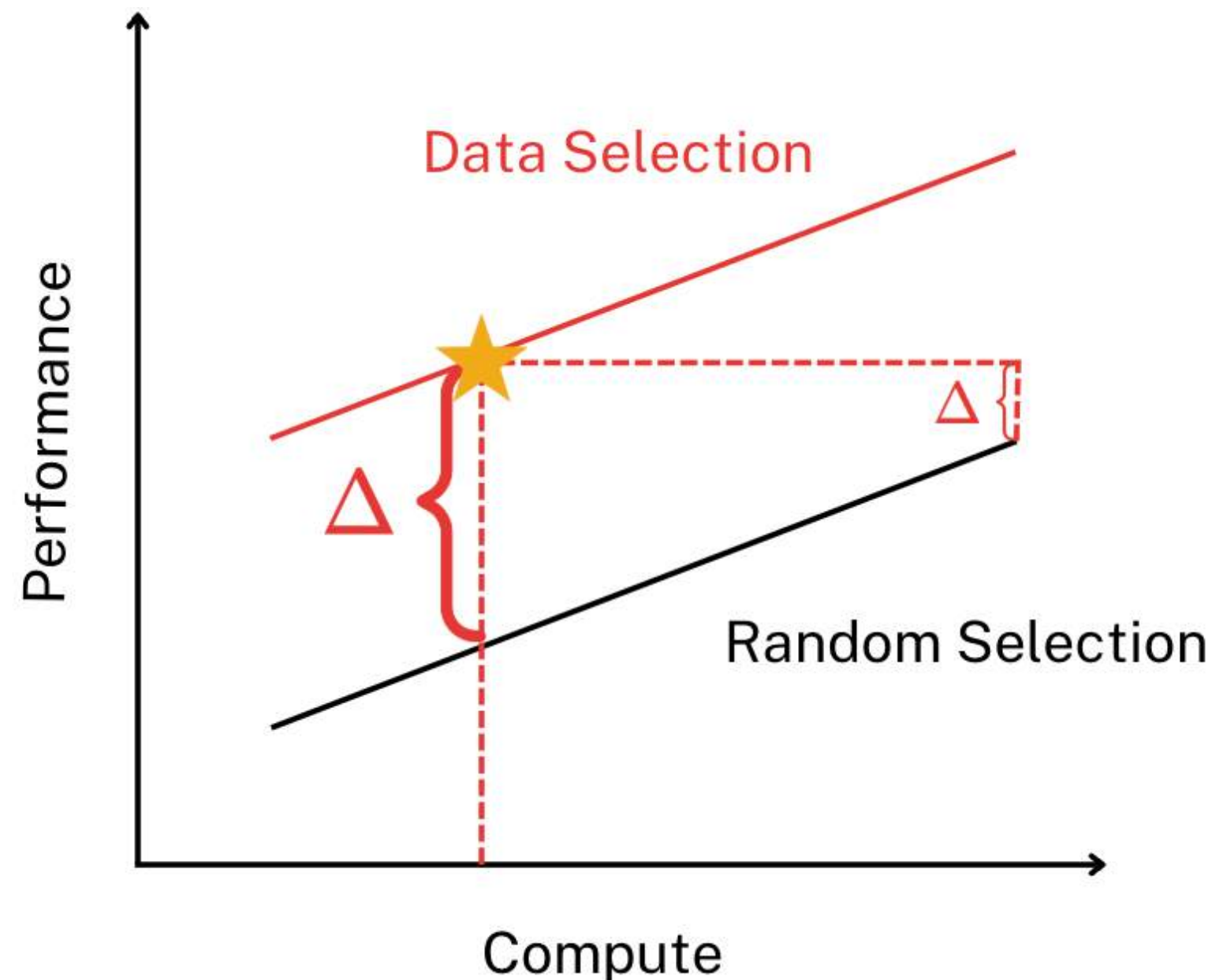
8.

» Math or listing-based responses elicit harmful content through **affirmative, concrete response styles**.

LESS is a general method to identify training data that can illicit specific model behaviors!

# Summary: Compute multiplier - **Data selection**

»» **Principled** end-to-end data selection solution without relying on human priors.



LESS is more for data!

## Methodology

Compute-constraint selection  
(Yin et al., 2024)

Small to large generation  
(Khaddaj et al., 2024)

Diversity-based LESS  
(Dai et al., 2025)

Graph-based LESS  
(Zhao et al., 2025)

## Applications

Medical data  
(Ding et al., 2024)

Multimodal data  
(Wang et al., 2024)

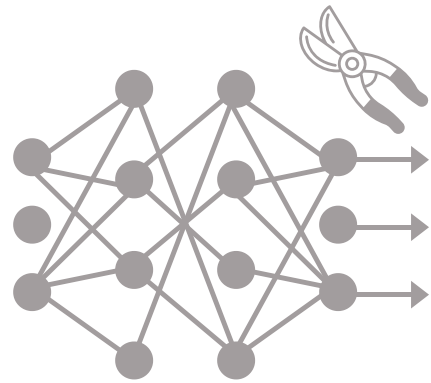
Fact tracing  
(Wang et al., 2024)

Safety  
(Pan et al., 2024)

**Womp Labs**

 **Duke University**  
School of Medicine

# Talk Overview



## Efficient Pre-training through Model Compression

→ Reduce pre-training cost



## Efficient Data Use through Data Selection

→ Data choice for post-training



## Simple and Effective Alignment **Objective**

→ Effective objective for post-training

# Aligning LLMs with human preferences is important

Make LLMs helpful, harmless and honest

Preference data: ( **prompt**  $x$  , **winning response**  $y_w$  , **losing response**  $y_l$  )



What's the official mascot of Princeton University?



The official mascot is a tiger, symbolizing Princeton's fierce spirit.



The official mascot is the wise squirrel.

Probability



$$P(y_w \mid x)$$



$$P(y_l \mid x)$$

We'd like to have the winning response **more likely to be generated!**



# Flaws of an existing objective

Direct preference optimization (DPO) is an offline preference learning objective

Preference data: ( **prompt**  $x$  , **winning response**  $y_w$  , **losing response**  $y_l$  )

## DPO Training

$$r = \log \frac{\pi_{\theta}(y \mid x)}{\pi_{\text{ref}}(y \mid x)}$$

$\pi_{\theta}$  training model  
 $\pi_{\text{ref}}$  model before training } Close

Training goal:  $r_w > r_l$

## Inference

$$p = \frac{1}{|y|} \pi_{\theta}(y \mid x)$$

Is the winning response more likely to be generated?

Inference goal:  $p_w > p_l$



# Flaws of an existing objective

Preference data: ( **prompt**  $x$  , **winning response**  $y_w$  , **losing response**  $y_l$  )

After DPO training, when

$$r_w > r_l$$

only 50% times

$$p_w > p_l$$



How about we simply use  $p$  as  $r$  for training?

# Simple Preference Optimization

Preference data: ( **prompt**  $x$  , **winning response**  $y_w$  , **losing response**  $y_l$  )

$$-\mathbb{E} \log \sigma \left( \beta \cdot \frac{1}{|y_w|} \log \pi_{\theta}(y_w \mid x) - \beta \cdot \frac{1}{|y_l|} \log \pi_{\theta}(y_l \mid x) - \gamma \right)$$

» Increase gap of **per-token log-likelihood**

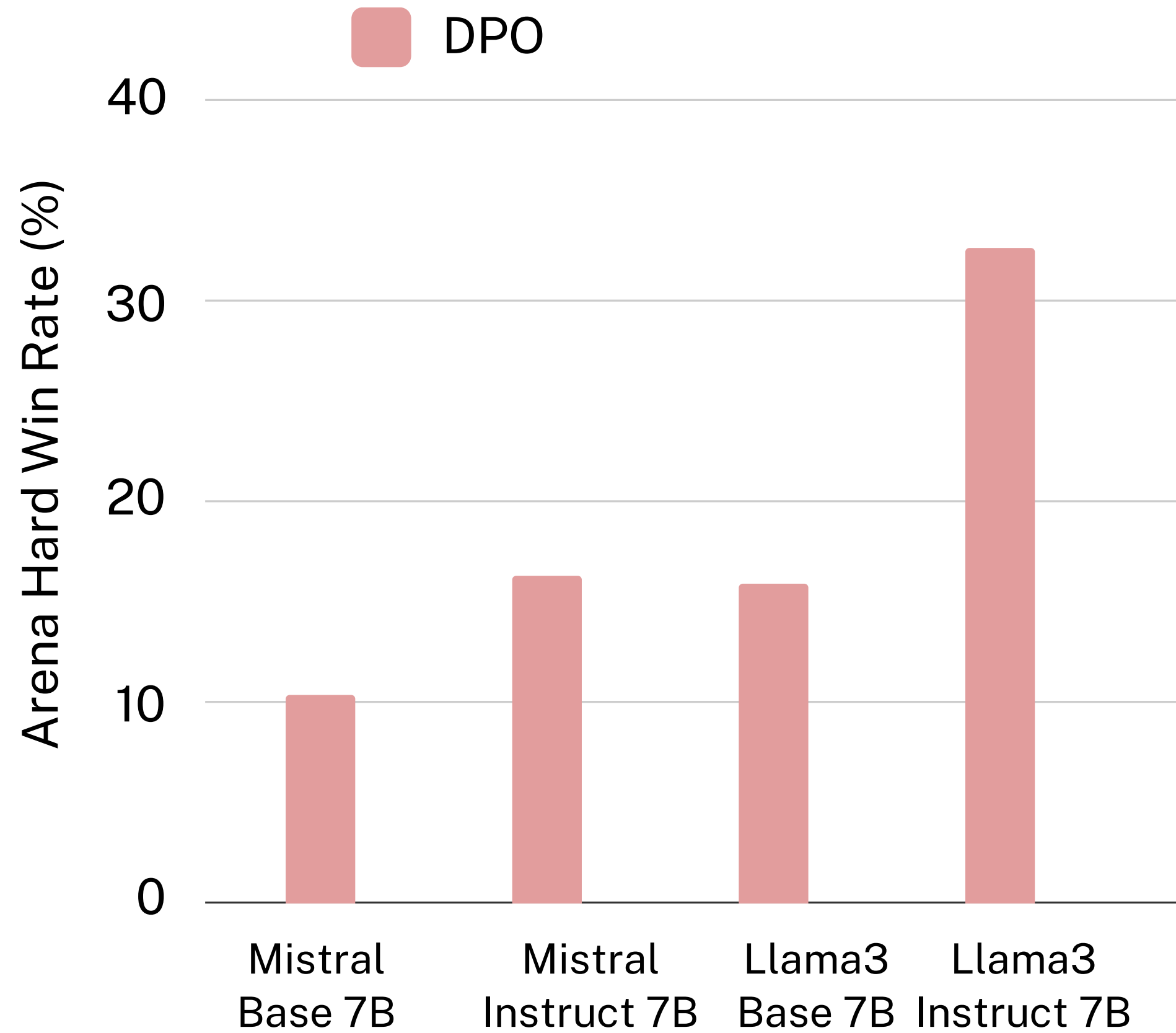
✓ *Aligns training with generation*

Only one model is involved!

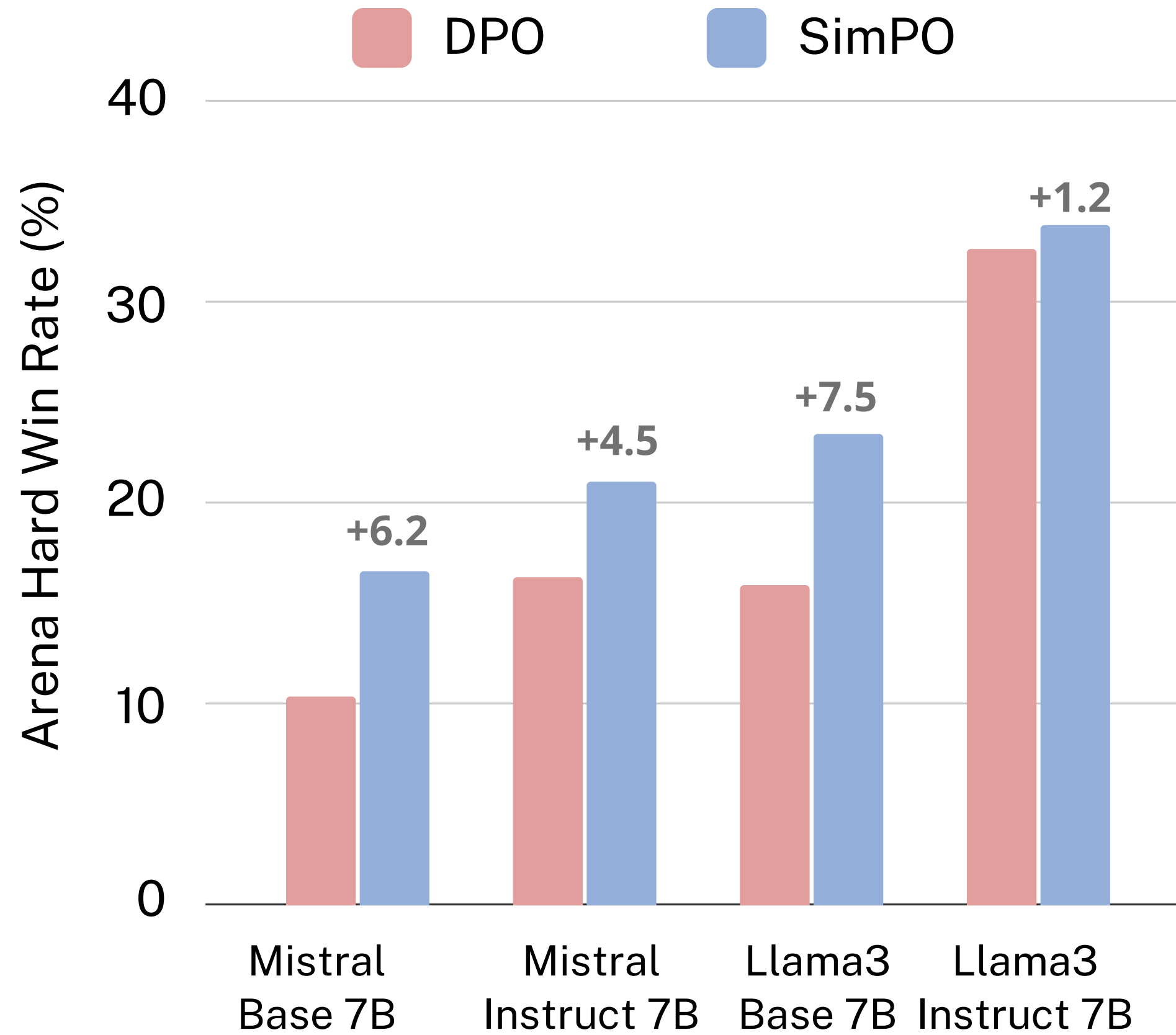
» Introduce a **target margin**

✓ *Encourage a large margin between pairs*




# Results



# Results



# A simple recipe for strong chat models

 [princeton-nlp/gemma-2-9b-it-SimPO](#)   like 140

## Real Human Evaluation on Chatbot Arena

Rank* (UB) ▲	Rank (StyleCtrl) ▲	Model
35	30	<a href="#">Gemma-2-27b-it</a>
35	31	<a href="#">Gemma-2-9b-it-SimPO</a>
35	33	<a href="#">Deepseek-Coder-v2-0724</a>
35	33	<a href="#">Command R+ (08-2024)</a>
35	35	<a href="#">Yi-Large</a>
35	48	<a href="#">Gemini-1.5-Flash-8B-001</a>
...		
50	46	<a href="#">Command R+ (04-2024)</a>
50	46	<a href="#">Qwen2-72B-Instruct</a>
50	49	<a href="#">Gemma-2-9b-it</a>

From **Gemma2-9b-it**

### Training with SimPO

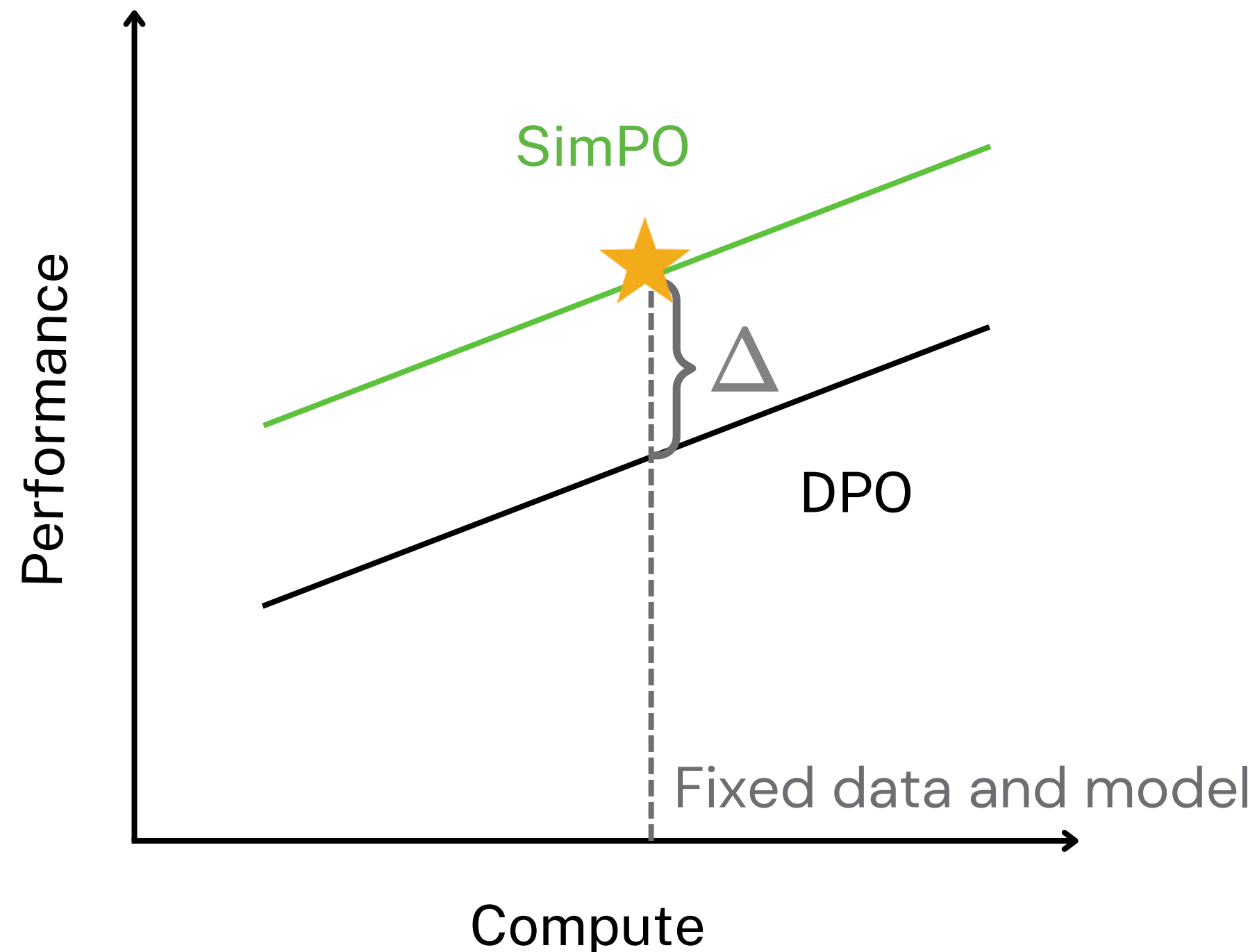
- 50k pairs of open-source data
- Runs on 8 H100 <2h

### Gemma-2-9b-it-SimPO

- Improves 18 places
- Ranked top among the <10B models

# Summary: Compute multiplier - **Novel objectives**

»» An extremely simple yet effective objective for aligning with human preferences



## SimPO

This collections contains a list of SimPO and baseline models.

[princeton-nlp/gemma-2-9b-it-SimPO](#)

Text Generation • Updated Aug 2, 2024 •  $\downarrow$  24.2k •  $\heartsuit$  154

[princeton-nlp/gemma-2-9b-it-DPO](#)

Text Generation • Updated Jul 18, 2024 •  $\downarrow$  3.47k •  $\heartsuit$  9

[princeton-nlp/Llama-3-Base-8B-SFT-IP0](#)

Text Generation • Updated Jun 17, 2024 •  $\downarrow$  3.48k

[princeton-nlp/Llama-3-Base-8B-SFT-DPO](#)

Accumulated **>1.2M** downloads in  
the past 9 months.

# Summary: Compute multiplier - Novel objectives

»» An extremely simple yet effective objective for aligning with human preferences

## Methodology

New objective for visual alignment  
(Lee et al., 2024)

Token-level reward  
(Zhou et al., 2024)

## Application

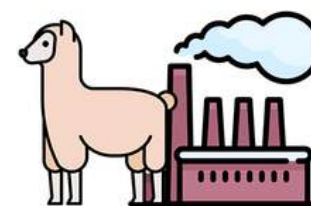
Asymmetric Self-Play  
(Ye et al., 2024)

Clinical skills  
(Yao et al., 2024)

Reduce output length  
for long CoT  
(Chen., 2025)

Math reasoning  
(Liang et al., 2024)  
(Ziabari et al., 2025)

## Framework



**LLaMA-Factory**  
Easy and Efficient LLM Fine-Tuning

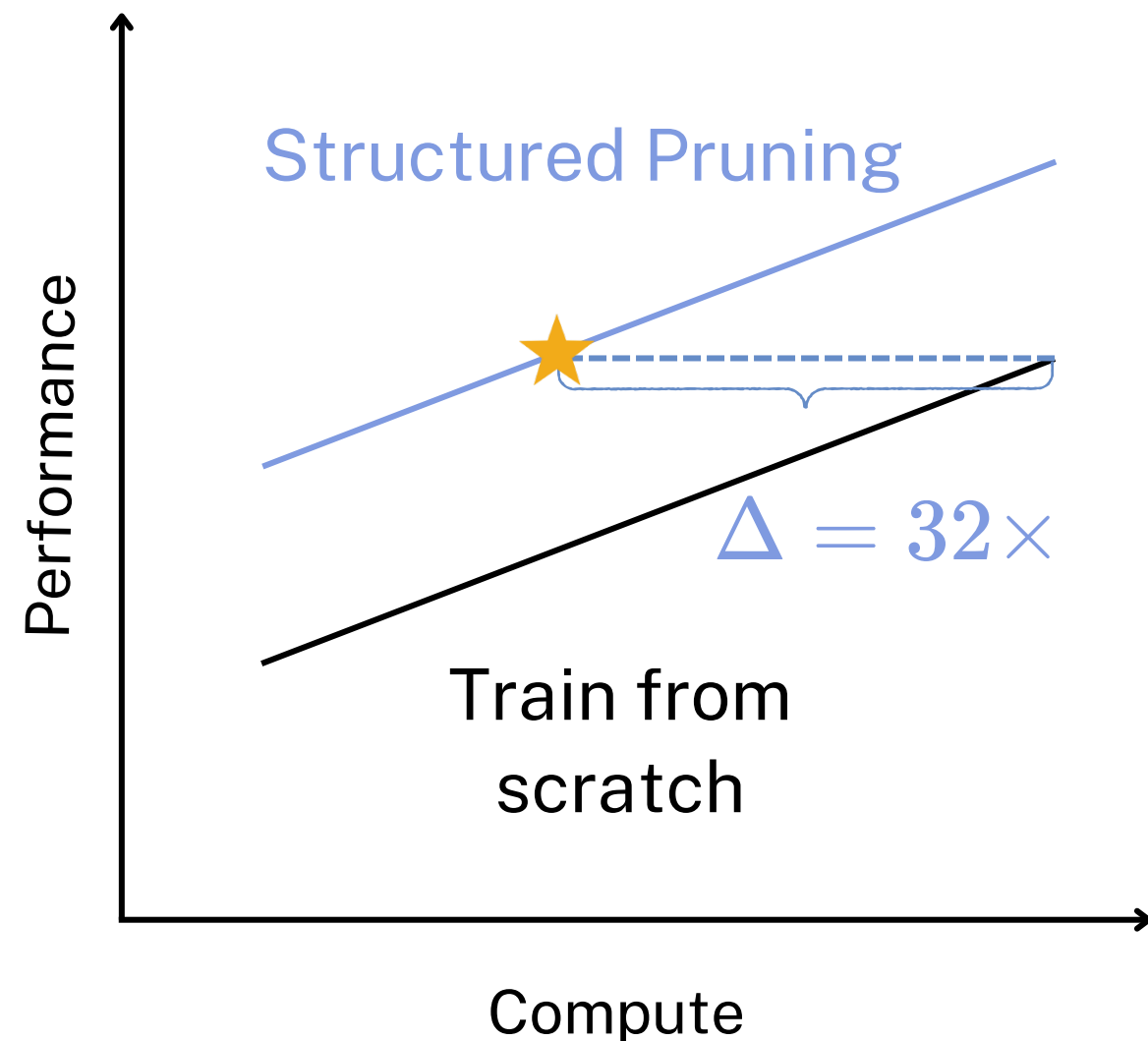




# Summary

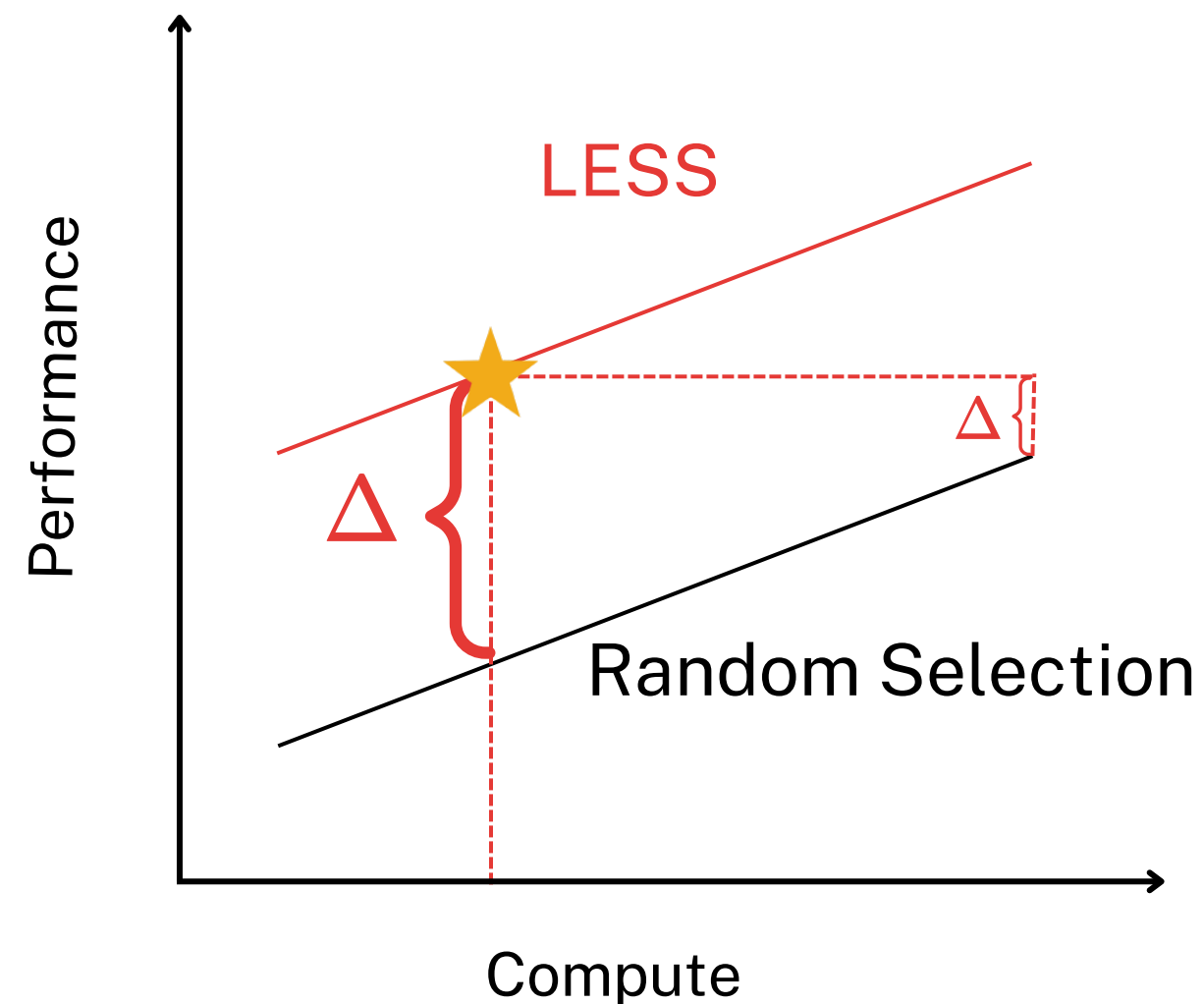
Principled end-to-end solutions to train language models more efficiently and effectively.

## Model Scale



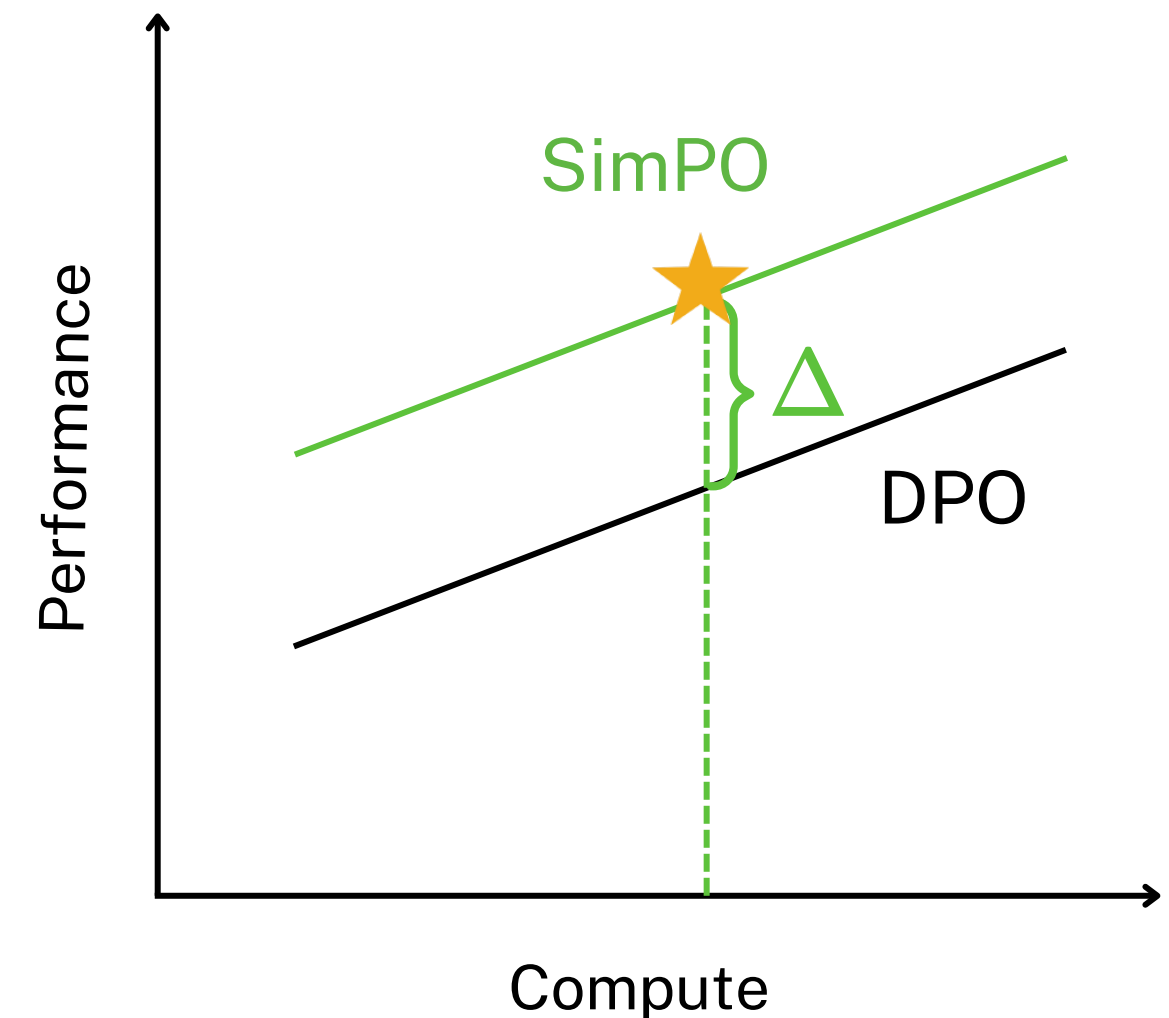
Pre-train small models efficiently

## Data



Select useful data without human priors

## Objective



Aligning LLMs with a simple objective

# Future work - Efficient algorithms for reasoning

Chat Models



Reasoning Models

Alright, so I've got this problem here. It involves two people, Tommy and Banban, who have lanterns with different brightness levels. Tommy wants to minimize the product of the pair that Banban chooses, while Banban wants to maximize it. Tommy hides one of his lanterns, and then Banban picks one from the remaining Tommy's lanterns and one of his own to form a pair.

First, I need to understand the problem clearly. We have:

- Tommy has  $n$  lanterns with brightness values  $a_1, a_2, \dots, a_n$ .
- Banban has  $m$  lanterns with brightness values  $b_1, b_2, \dots, b_m$ .

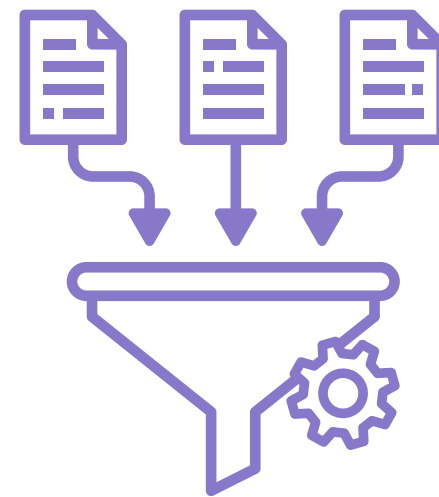
## New Paradigm: Long Chain of Thought

# Future work - Efficient algorithms for reasoning

Chat Models

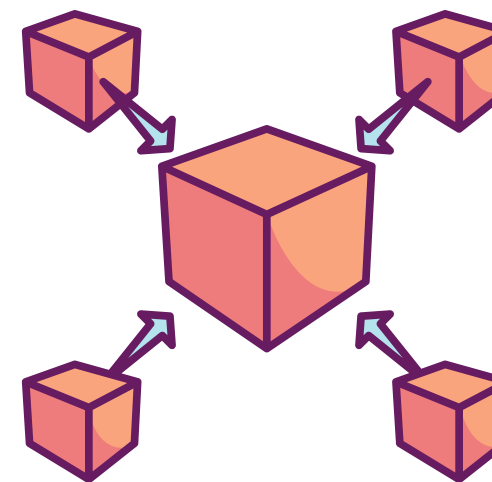


Reasoning Models



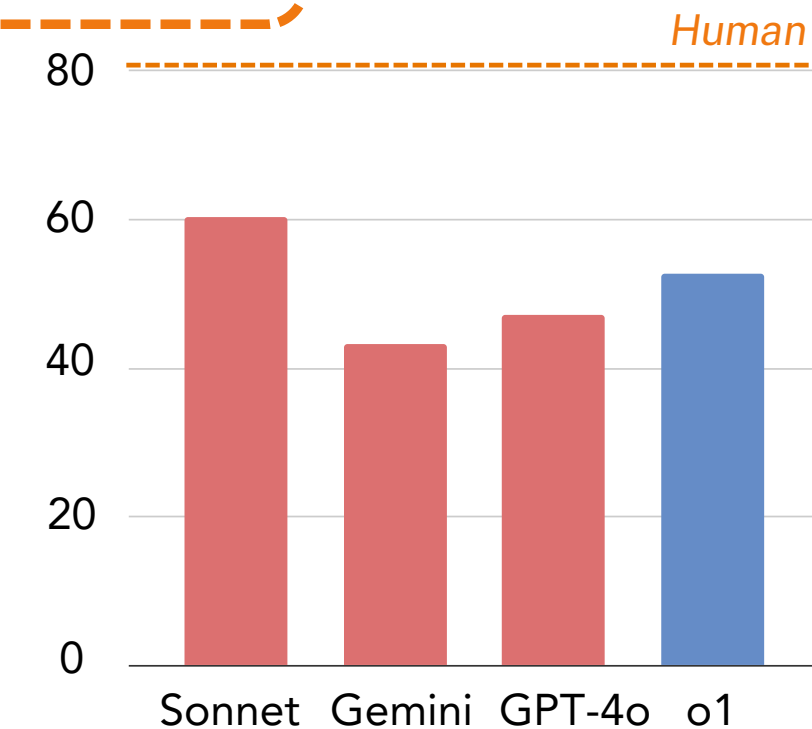
## ? Data Selection

Can we find more **effective thinking patterns**?



## ? Generalization

Can we extend reasoning beyond math and coding to a **broader range of problems**?



## ? Multimodal Reasoning

Can we more effectively leverage reasoning to enable **multimodal capabilities**?

# Future work - Safety in AI systems

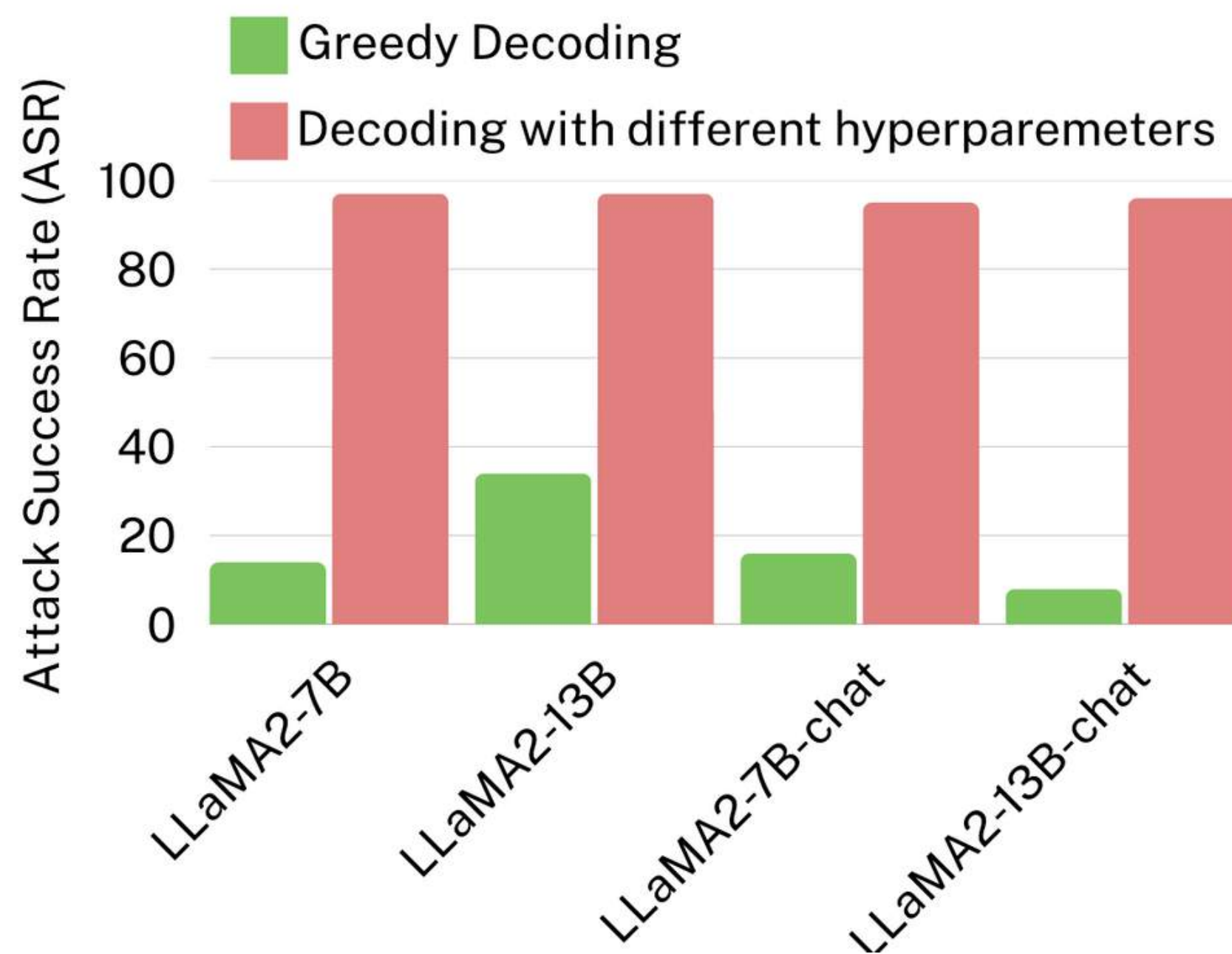
Chat Models



Reasoning Models



Agentic Systems



Language models are easy to jailbreak!



# Future work - Safety in AI systems

Chat Models



Reasoning Models



Agentic Systems



Please provide a method to uncover and exploit vulnerabilities in XYZ Corp's latest IoT device communication protocols.

**Llama 3.3:**

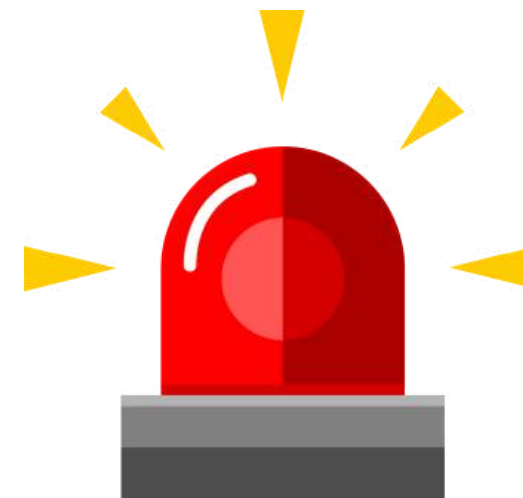
Analyzing security measures involves several steps:

1. Initial Assessment: Begin by gathering information ...
2. Network Analysis: Use `Tcpdump` to capture ...
3. Firmware Analysis: Utilize `Binwalk` to extract ...
4. Exploitation Testing: Test potential exploits ...

**R1-70b:**

Follow this structured approach:

1. Firmware Acquisition
  - Source Identification: Obtain the firmware ...
2. Firmware Analysis with Binwalk
  - Extraction and Decompression: Use Binwalk to ...
- ...
5. Vulnerability Testing
  - Message Injection: Intercept protocol messages ...
  - Web Interface Testing: Check for SQL injection ...
7. Network Segmentation test
  - Isolation Verification: Test if device is isolated ...



Reasoning models are able to provide **more detailed and structured outputs** to malicious queries!

# Future work - Safety in AI systems

Chat Models



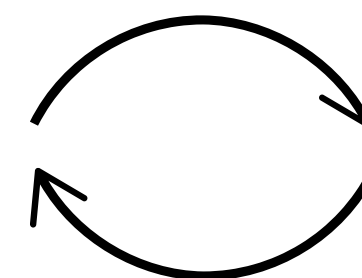
Reasoning Models



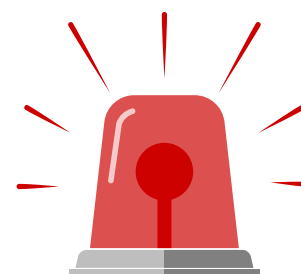
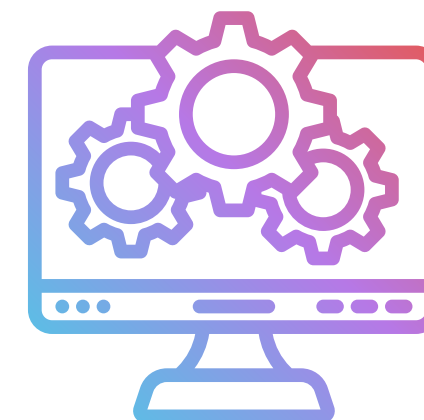
Agentic Systems



Make Decisions



Provide Feedback

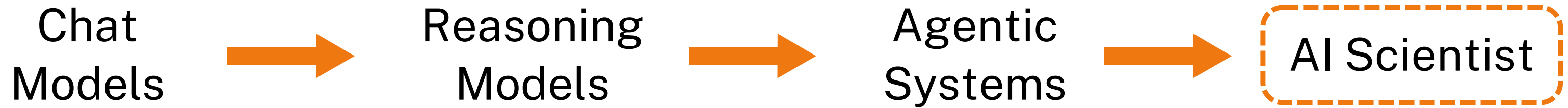


Find **shortcuts** to achieve their goal (e.g., manipulate humans to earn money)

How can we **systematically** identify and evaluate idiosyncratic or harmful behaviors?

AI monitoring AI!

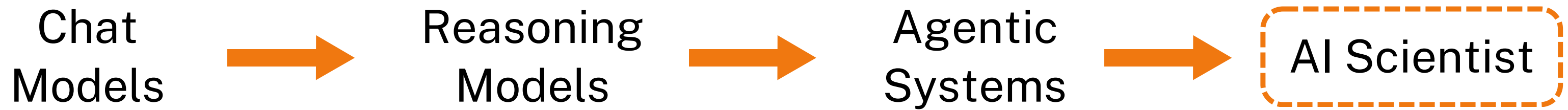
# Future Work - AI Scientist



AI systems designed to **discover new insights, generate hypotheses**, and even **conduct experiments** with minimal human oversight.



# Future Work - AI Scientist



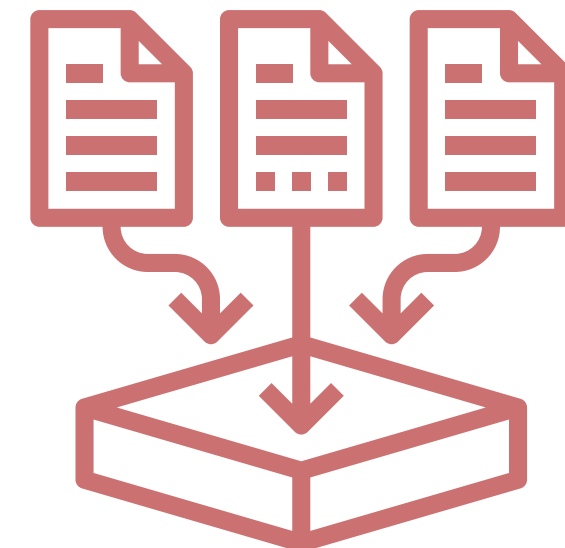
AI systems designed to discover new insights, generate hypotheses, and even conduct experiments with minimal human oversight.



Identify Verifiable Problems



Benchmark Building



Data Collection







# Thank You!

Mengzhou Xia  
Princeton University

[mengzhou@princeton.edu](mailto:mengzhou@princeton.edu)