The Best Instruction Tuning Data Are Those That Fit

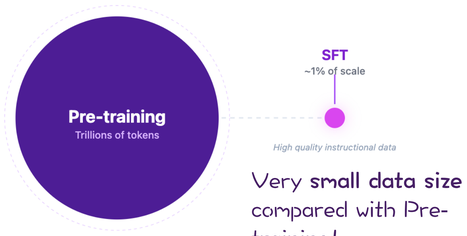
Dylan Zhang, Qirun Dai, Hao Peng

Reach out to: shizhuo2@Illinois.edu

NEURAL INFORMATION PROCESSING SYSTEMS

Data Scale in LLM Training

Supervised Fine-Tuning (SFT) uses just one percent of the data scale required for Pre-training.



- **Pre-training** — Massive ingestion of raw internet data to learn general language patterns, reasoning, and world knowledge.
- **Supervised Fine-Tuning** — A tiny, curated subset of demonstration data used to guide the model's behavior and output format.

MOTIVATION & HYPOTHESIS

⚠️ The Problem: Off-Policy Data

- Causes **Catastrophic Forgetting**
- Results in **Inefficient Learning**

Key Insight

Research demonstrates that **On-Policy-ness** is a critical success factor for both Preference Optimization and Reinforcement Learning.

HYPOTHESIS

"Quality is not enough!"

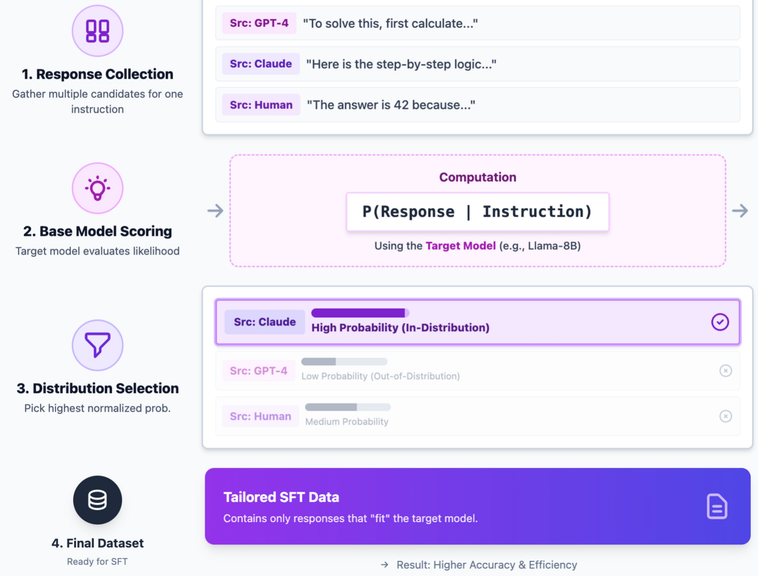
Standard SFT → 🎯 Match Pre-trained Distribution

SFT must supervise LMs on targets that align with their internal distribution.

🧠 FORMULATE THIS AS A DATA SELECTION PROBLEM

How GRAPE Works

Aligning Instruction-Tuning Data to the Target Model's Distribution

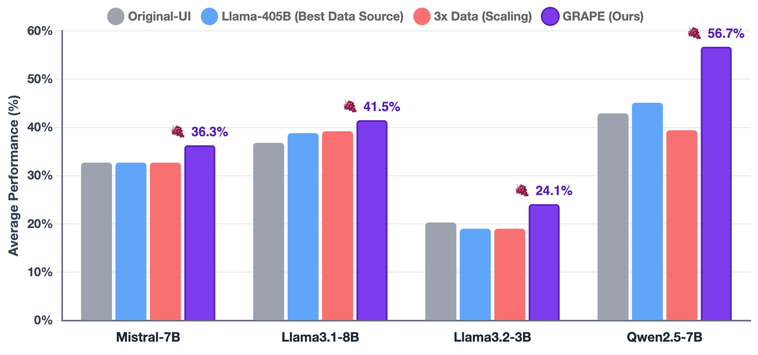


EXP1: VERIFIABLE COT TRAINING

UltraInteract-SFT Prompts → Sample CoTs From Various Model → Verify By Answer Correctness → Apply GRAPE

GRAPE Performance Advantage

Average Performance on UltraInteract-SFT Benchmarks



Key Insights:

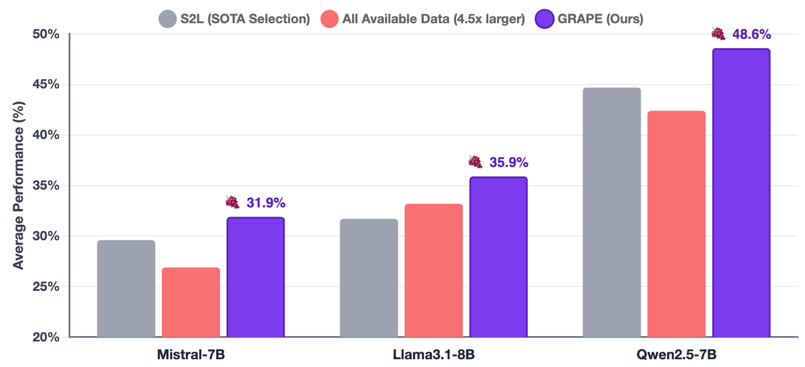
- **Superior Performance:** GRAPE outperforms all baselines across every model family.
- **Qwen2.5 Gains:** 17.3% gain over scaling and 13.8% over the original dataset.
- **Efficiency:** GRAPE consistently improves performance where scaling data often fails.

EXP2: BETTER MODELS FROM EXISTING DATA

Olmo Tulu Post-training+ Prompts → Subset Prompts with Multiple Responses → Form Candidate Set Using Existing Responses → Apply GRAPE

GRAPE Real-World Effectiveness

Performance on Tulu-3 / Olmo-2 Data Mixture



Key Insights:

- **Beats Scaling:** GRAPE outperforms training on "All Available Data" (1.58M instances) using only 22% of the data volume (350k instances).
- **Surpasses SOTA:** GRAPE consistently exceeds state-of-the-art selection methods like S2L (SmallToLarge) across all model families.
- **Qwen2.5 Boost:** Achieves a massive +6.2% gain over the "All Available Data" baseline on Qwen2.5-7B.

SIMPLE → SCALABLE!

DATA SELECTION METHOD	+ TRAINING COST	PER-SAMPLE COMPUTATION
GRAPE (Ours)	0	$N \times F_\theta$
LESS	$C(\theta_{lora}, D_{warmup}, T)$	$3T \times N \times F_\theta$
Embedding-based	0	$N \times F_\theta$
S2L	$C(\theta_{ref}, D, T)$	$T \times N \times F_{\theta ref}$

Key Takeaway:

GRAPE needs **no extra training** and only $N \times F_\theta$ per sample, while gradient or trajectory methods rerun training loops.

N: training dataset size | T: number of updates | F_θ: cost of one forward pass

What's next?

Supervision shall fit the base model for SFT

- **Data (GRAPE)** ✅
- **Update Objective** ⏳

Check out our demo page at: https://dylanzsz.github.io/LIME/

Personal Page

Project Page

Email: shizhuo2@Illinois.edu **X / Twitter:** https://x.com/dylan_works_