Breaking the Resource Monopoly: LLM Post-Training and Serving with Modest Data and Compute

Jiaxin Huang

Computer Science and Engineering
Washington University in St. Louis
jiaxinh@wustl.edu

Abstract

Frontier large language models are increasingly powerful though many of them are trained from vast proprietary data and intensive computes, raising barriers for academic labs and smaller institutions for exploration and improvement. In this talk, I will present a unified research agenda for breaking the resource monopoly in both post-training and serving. On the training side, I will describe label-free and even zero-data post-training pipelines that let models curate their own reasoning supervision. On the serving side, I will show how cost-aware inference can enable adaptive test-time scaling to be more efficient. Together, these components form a practical LLM system using modest data and compute resources.

Part I: Language Model Post-Training via Label-Free and Zero-Data Curation

The first part covers how to conduct language model post-training on challenging reasoning questions by replacing large human-labeled corpora with model-curated training signals. **First**, our previous work LMSI (Huang et al. 2023) introduces a generate–filter–refine pipeline that selects high-confidence rationales as pseudo-labels, enabling substantial gains in reasoning without any ground-truth solutions. **Building on this**, R-Zero (Huang et al. 2025b) eliminates even the need for manually creating reasoning questions: a Challenger model constructs increasingly difficult problems, while a Solver model learns to solve them. **Furthermore**, to ensure we measure the right capabilities, CrossWordBench (Leng et al. 2025a) introduces controllable puzzle generation for fine-grained tests of compositional reasoning, allowing systematic comparison across models as data curation strategies improve. Finally, I will touch on how calibrated signals can guide which pseudo-labels to trust, which ensures that models learn when to trust their own generated supervision.

Part II: Trustworthy and Cost-aware Language Model Serving

The second part covers language model serving from the aspect of reliability and inference efficiency. **First**, I will introduce our PPO-M and PPO-C (Leng et al. 2025b) which shows that standard RLHF often amplifies verbalized overconfidence, and this is driven by biases in reward modeling. We propose calibrated RLHF variants (PPO-M, PPO-C) that align verbalized confidence with accuracy without sacrificing language model capability on general tasks. This makes post-trained models safer for deployment and better suited to serve as teachers in self-curated pipelines. **What's more**, the same calibration principles can largely benefit cost-aware inference. We design an efficient test-time scaling approach (Huang et al. 2025a) to adaptively allocate sampling budgets between easy and difficult queries. Our method largely reduces the computational cost needed to solve a large bulk of questions while maintaining the accuracy. **Complementarily**, our PosS (Huang et al. 2025c) introduces specialized draft models for speculative decoding to improve language model inference speed. These various components can be integrated into an efficient LLM serving system that routes queries by predicted difficulty, uses the smallest adequate model when possible, and employs calibrated test-time scaling only where needed.

References

Huang, C.; Huang, L.; Leng, J.; Liu, J.; and Huang, J. 2025a. Efficient Test-Time Scaling via Self-Calibration. *ArXiv*, abs/2503.00031.

Huang, C.; Yu, W.; Wang, X.; Zhang, H.; Li, Z.; Li, R.; Huang, J.; Mi, H.; and Yu, D. 2025b. R-Zero: Self-Evolving Reasoning LLM from Zero Data. *ArXiv*, abs/2508.05004.

Huang, J.; Gu, S. S.; Hou, L.; Wu, Y.; Wang, X.; Yu, H.; and Han, J. 2023. Large Language Models Can Self-Improve. In *EMNLP*.

Huang, L.; Huang, C.; Leng, J.; Huang, D.; and Huang, J. 2025c. POSS: Position Specialist Generates Better Draft for Speculative Decoding. *ArXiv*, abs/2506.03566.

Leng, J.; Huang, C.; Huang, L.; Lin, B. Y.; Cohen, W. W.; Wang, H.; and Huang, J. 2025a. CrossWordBench: Evaluating the Reasoning Capabilities of LLMs and LVLMs with Controllable Puzzle Generation. In *COLM*.

Leng, J.; Huang, C.; Zhu, B.; and Huang, J. 2025b. Taming Overconfidence in LLMs: Reward Calibration in RLHF. In *ICLR*.