High-quality And Efficient Post-Training Of Large Models Based On Synthetic Data

Haoyu Wang
Beijing University of Posts and Telecommunications

Augest, 2024

Abstract

This research proposal focuses on addressing the challenges associated with aligning large models using synthetic data within the domain of natural language processing. As the proliferation of large models intensifies, the scarcity of high-quality data becomes pronounced, and synthetic data emerges as a potential solution. However, current approaches encounter issues such as the reliance on human-dependent data curation, difficulties in quality inspection, and suboptimal data selection and formulation.

The research centres on three key questions: large-scale multi-agent LLMs data curation, data quality inspection based on Superalignment, and synthetic data formula. The aim is to enhance the autonomy of data generation, establish an efficient inspection mechanism, and optimize data arrangement.

The study is planned over a period of four years, encompassing literature review, method development and experimentation, and finalization for dissemination. Anticipated outcomes include novel and effective approaches, contributing to the improvement of large model alignment and providing valuable guidelines for the field.

1 Introduction

In the contemporary field of natural language processing, achieving high-quality alignment has emerged as the crux for enhancing model performance and reliability. The rapid advancement of large models has led to a depletion of high-quality data (Villalobos et al., 2024). Correspondingly, obtaining high-quality alignment data also faces similar predicaments. Human annotators have traditionally been the primary source of high-quality alignment data (OpenAI et al., 2024). A synthetic data method has been proposed to alleviate this issue. This approach is predicated on a powerful large language model (LLM) itself generating alignment data (Ding et al., 2023; Peng et al., 2023; Taori et al., 2023), which plays a vital role in eliciting the instruction-following ability of LLMs and aligning the model behaviour with human preferences. It efficiently and economically acquires a substantial amount of aligned data, which is beneficial for expanding the model's capability boundaries (C. Xu et al., 2023) and facilitating adaptation to downstream tasks (Long et al., 2024; Wang et al., 2024). Currently, this method has been widely adopted in the training of leading open-source large models (Llama Team, 2024; Nvidia et al., 2024; Yang et al., 2024), signifying its promising prospects. Nevertheless, the current method is still in its nascent stage, and numerous problems await resolution. Regarding the data curation method, it is currently primarily based on prompt engineering to construct single or multiple agents for building the manufacturing pipeline (Ding et al., 2023; Wang et al., 2024; C. Xu et al., 2023), demanding the design of a specific framework for specific tasks. Moreover, in the face of massive synthetic data, especially in multilingual (Singh et al., 2024; Wang et al., 2024), long-context (Bai et al., 2024; Y. Chen et al., 2024), and multi-modal (Y. Liu et al., 2024) domains, enhancing human quality inspection poses a significant challenge.

Furthermore, using synthetic data for model training without appropriate selection can result in model collapse (Shumailov et al., 2024), highlighting the significance of data selection and filtering to ensure favourable distribution and overall quality. This constitutes a promising area warranting further exploration.

2 Survey of Background Literature

2.1 Data Curation

Currently, synthetic data has been extensively utilized in both the supervised fine-tuning (SFT) (Llama Team, 2024; Nvidia et al., 2024; Yang et al., 2024) and the reinforcement learning from human feedback (RLHF) (Cui et al., 2024; Lee et al., 2023) phases.

In the SFT domain, there are mainly two directions: large-to-small and small-to-large (Long et al., 2024). The large-to-small approach involves training small language models (Abdin et al., 2024; Hu et al., 2024) or fine-tuning task-specific large language models. Through a meticulously designed data processing pipeline, the large language model leverages seed data or its own capability to generate specific SFT data to enhance the relevant functions. Conversely, the small-to-large approach represents the training of more powerful large language models or self-improvement. It requires the design of a meta method for self-evolution (C. Xu et al., 2023) or involves generating a large quantity and conducting searches within the generated results (Luo et al., 2024; Tian et al., 2024; C. Xu et al., 2023).

In the RLHF field, the large language model is employed to substitute the human annotator in the rating phase (Cui et al., 2024; Lee et al., 2023; McAleese et al., 2024). This area is on an upward trajectory and awaits further exploration.

2.2 Data Quality Inspection

With the mass production of synthetic data, ensuring its correctness has become an inescapable issue. Currently, the mainstream approach is manual inspection (Y. Liu et al., 2024; Singh et al., 2024; Wang et al., 2024), but this method is costly. Even more challenging is that sometimes it is arduous for human annotators to handle certain data, such as extremely long data (exceeding 128k), multilingual data (Wang et al., 2024), and multi-modal data (Y. Liu et al., 2024). There is an imperative need to establish a quality inspection mechanism to enhance the quality and efficiency of the inspection process.

2.3 Data Selection and Formula

After obtaining high-quality correct data points, it is indispensable to select the most valuable data to reduce the training cost and enhance performance (W. Liu et al., 2024). Currently, a mainstream method for evaluating abstract indicators relies on the prompt of LLMs (Ding et al., 2023; W. Liu et al., 2024), which is highly dependent on the fundamental ability of LLMs and has a propensity for their own contents (Panickssery et al., 2024). Hence, how to design a comprehensive indicator scheme and non-LLM indicators (Ding et al., 2023; Pillutla et al., 2021) to reduce the cost of selection and improve the overall quality is a crucial question.

Besides, the distribution of synthetic data exhibits a disparity with real data, such as information loss at the tail of the distribution (X. Xu et al., 2024) and augmented prejudice and discrimination (Shumailov et al., 2024). This can be mitigated by manually designing a data formula to modify the distribution (Fu et al., 2024) and integrating real data (Shumailov et al., 2024). However, there remains a deficiency in the methodology for composing data rather than merely an experimental plan, and this area demands further exploration.

3 Research Questions

1. **Large Scale Multi-Agent LLMs Data Curation**:
   Current methods mainly focus on building a data production pipeline using human prior knowledge, while the potential of vast amounts of models automatically generating data (Li et al., 2024) awaits exploration. A large amount of inference has been approved that has ability to improve the coverage and ability of LLMs based on good checkers (Brown et al., 2024). The development of infrastructure for multi-agent (W. Chen et al., 2024) lays a foundation for large-scale multi-agent self-organized inference. Further exploration is necessary on the autonomous generation, selection and iteration of multiple models based on game theory and sociology.

2. **Data quality inspection enhancement mechanism based on Superalignment**:

Correct data is the basis for a high-quality dataset. With the expansion in the scale of synthetic data, manual checking of each data point is unaffordable. There is an urgent need to establish an efficient data quality enhancement mechanism. Superalignment (Burns et al., 2023) shows potential in addressing this problem (McAleese et al., 2024). Through weak supervision, LLMs can independently correct mistakes. With the assistance of LLM, quality inspection is expected to complement and even surpass human-only quality inspection (Hosking et al., 2024).

3. **Data Formula With Synthetic Data**:
   Identifying suitable indicators to represent the quality of a group of data is challenging. Direct utilization of vanilla synthetic data can lead to model collapse after several generations (Shumailov et al., 2024). How to arrange data in different training stages (Hu et al., 2024) and in data distribution (Fu et al., 2024) significantly influences the final performance of LLM. With a well-designed data formula in both the composition and training stages, the side effect of distribution shift can be minimized. For highly customized features, LLM can even enhance certain weak abilities in real scenarios through deliberately designed synthetic data.

4 Research Aim And Objectives

4.1 Research Aim

To develop effective and efficient strategies for aligning large models using synthetic data, ensuring high quality and performance in various applications.

4.2 Objectives

1. Explore innovative approaches for large-scale multi-agent LLMs data curation to enhance the autonomy and self-organization of model generation and iteration.

2. Design and implement an advanced data quality inspection mechanism based on Superalignment to improve the accuracy and reliability of synthetic data.

3. Establish a comprehensive data formula framework for synthetic data to optimize the data arrangement in different training stages and minimize distribution shift effects.

4. Evaluate the performance and effectiveness of the proposed strategies through extensive experiments and comparisons with existing methods.

5. Analyze the impact of synthetic data on the customization and improvement of specific abilities of large models in real-world scenarios.

6. Provide guidelines and best practices for the application of synthetic data in large model alignment to promote the development and practical utilization of advanced technologies.

5 Research Method

1. **Literature Review**: Thoroughly collect and analyze the existing literature on multi-agent model data management, Superalignment data quality inspection, synthetic data formulas, and large model alignment to understand the cutting-edge achievements and existing problems in related fields.

2. **Experimental Research**: Design and conduct a series of experiments to verify the effectiveness of the proposed strategies and methods.

3. **Model Development and Training**: Construct new multi-agent models and data management infrastructures to explore the possibility of autonomous generation and iteration. Develop a quality inspection module based on Superalignment and integrate it into the data processing flow for testing.

4. **Data Analysis**: Conduct detailed analyses of the data generated from experiments, including the accuracy, reliability, and distribution characteristics of the data. For instance, analyze the distribution changes of synthetic data in different training stages to evaluate the effect of the data formula.

6 Expect Results

Firstly, in the area of large-scale multi-agent data curation, it is expected to discover novel and effective methods that significantly enhance the autonomy and self-organization of model generation and iteration. This could lead to more efficient and flexible data curation processes, reducing the reliance on human prior knowledge and enabling models to adapt and evolve more autonomously.

Secondly, for the data quality inspection enhancement mechanism based on Superalignment, it is anticipated to establish an efficient and accurate inspection mechanism that outperforms traditional methods and manual inspection in terms of accuracy and reliability. This would ensure the quality of synthetic data and improve the performance and trustworthiness of large models.

In terms of the data formula with synthetic data, it is expected to develop a comprehensive and effective data formula framework that optimizes the data arrangement in different training stages and minimizes the side effects of distribution shifts. This would contribute to a more stable and improved final performance of LLM.

7 Research Plan

Year 1

- Carry out an in-depth literature review regarding synthetic data formulas, multi-agent systems, and Superalignment.
- Identify research voids and formulate precise research inquiries.
- Fortify the foundation of mathematics and the history of machine learning.

Year 2

- Develop initial methods for large-scale multi-agent LLMs data curation and data quality inspection mechanisms.
- Carry out small-scale experiments and analyze the preliminary results.

Year 3

- Optimize the methods based on the Year 2 results.
- Conduct large-scale experiments and evaluate the performance.
- Analyze the impact of synthetic data on model customization.

Year 4

- Finalize the research and write the dissertation.
- Prepare for the defense and disseminate the research results.

References

Abdin, M., Jacobs, S. A., Awan, A. A., Aneja, J., Awadallah, A., Awadalla, H., Bach, N., Bahree, A., Bakhtiari, A., Bao, J., Behl, H., Benhaim, A., Bilenko, M., Bjorck, J., Bubeck, S., Cai, Q., Cai, M., Mendes, C. C. T., Chen, W., . . . Zhou, X. (2024). Phi-3 technical report: A highly capable language model locally on your phone. https://arxiv.org/abs/2404.14219

Bai, Y., Lv, X., Zhang, J., He, Y., Qi, J., Hou, L., Tang, J., Dong, Y., & Li, J. (2024). Longalign: A recipe for long context alignment of large language models. https://arxiv.org/abs/2401.18058

Brown, B., Juravsky, J., Ehrlich, R., Clark, R., Le, Q. V., Ré, C., & Mirhoseini, A. (2024). Large language monkeys: Scaling inference compute with repeated sampling. https://arxiv.org/abs/2407.21787

Burns, C., Izmailov, P., Kirchner, J. H., Baker, B., Gao, L., Aschenbrenner, L., Chen, Y., Ecoffet, A., Joglekar, M., Leike, J., Sutskever, I., & Wu, J. (2023). Weak-to-strong generalization: Eliciting strong capabilities with weak supervision. https://arxiv.org/abs/2312.09390

Chen, W., You, Z., Li, R., Guan, Y., Qian, C., Zhao, C., Yang, C., Xie, R., Liu, Z., & Sun, M. (2024). Internet of agents: Weaving a web of heterogeneous agents for collaborative intelligence. https://arxiv.org/abs/2407.07061

Chen, Y., Qian, S., Tang, H., Lai, X., Liu, Z., Han, S., & Jia, J. (2024). Longlora: Efficient fine-tuning of long-context large language models. https://arxiv.org/abs/2309.12307

Cui, G., Yuan, L., Ding, N., Yao, G., He, B., Zhu, W., Ni, Y., Xie, G., Xie, R., Lin, Y., Liu, Z., & Sun, M. (2024). Ultrafeedback: Boosting language models with scaled ai feedback. https://arxiv.org/abs/2310.01377

Ding, N., Chen, Y., Xu, B., Qin, Y., Hu, S., Liu, Z., Sun, M., & Zhou, B. (2023). Enhancing chat language models by scaling high-quality instructional conversations. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. https://aclanthology.org/2023.emnlp-main.183

Fu, Y., Panda, R., Niu, X., Yue, X., Hajishirzi, H., Kim, Y., & Peng, H. (2024). Data engineering for scaling language models to 128k context. https://arxiv.org/abs/2402.10171

Hosking, T., Blunsom, P., & Bartolo, M. (2024). Human feedback is not gold standard. https://arxiv.org/abs/2309.16349

Hu, S., Tu, Y., Han, X., He, C., Cui, G., Long, X., Zheng, Z., Fang, Y., Huang, Y., Zhao, W., Zhang, X., Thai, Z. L., Zhang, K., Wang, C., Yao, Y., Zhao, C., Zhou, J., Cai, J., Zhai, Z., … Sun, M. (2024). Minicpm: Unveiling the potential of small language models with scalable training strategies. https://arxiv.org/abs/2404.06395

Lee, H., Phatale, S., Mansoor, H., Mesnard, T., Ferret, J., Lu, K., Bishop, C., Hall, E., Carbune, V., Rastogi, A., & Prakash, S. (2023). Rlaif: Scaling reinforcement learning from human feedback with ai feedback. https://arxiv.org/abs/2309.00267

Li, J., Zhang, Q., Yu, Y., Fu, Q., & Ye, D. (2024). More agents is all you need. https://arxiv.org/abs/2402.05120

Liu, W., Zeng, W., He, K., Jiang, Y., & He, J. (2024). What makes good data for alignment? a comprehensive study of automatic data selection in instruction tuning. https://arxiv.org/abs/2312.15685

Liu, Y., Cao, Y., Gao, Z., Wang, W., Chen, Z., Wang, W., Tian, H., Lu, L., Zhu, X., Lu, T., Qiao, Y., & Dai, J. (2024). Mminstruct: A high-quality multi-modal instruction tuning dataset with extensive diversity. https://arxiv.org/abs/2407.15838

Llama Team, A. @. M. (2024). The llama 3 herd of models. https://ai.meta.com/research/publications/the-llama-3-herd-of-models/

Long, L., Wang, R., Xiao, R., Zhao, J., Ding, X., Chen, G., & Wang, H. (2024). On llms-driven synthetic data generation, curation, and evaluation: A survey. https://arxiv.org/abs/2406.15126

Luo, H., Sun, Q., Xu, C., Zhao, P., Lin, Q., Lou, J., Chen, S., Tang, Y., & Chen, W. (2024). Arena learning: Build data flywheel for llms post-training via simulated chatbot arena. https://arxiv.org/abs/2407.10627

McAleese, N., Pokorny, R. M., Uribe, J. F. C., Nitishinskaya, E., Trebacz, M., & Leike, J. (2024). Llm critics help catch llm bugs. https://arxiv.org/abs/2407.00215

Nvidia, : Adler, B., Agarwal, N., Aithal, A., Anh, D. H., Bhattacharya, P., Brundyn, A., Casper, J., Catanzaro, B., Clay, S., Cohen, J., Das, S., Dattagupta, A., Delalleau, O., Derczynski, L., Dong, Y., Egert, D., Evans, E., ... Zhu, C. (2024). Nemotron-4 340b technical report. https://arxiv.org/abs/2406.11704

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., ... Zoph, B. (2024). Gpt-4 technical report. https://arxiv.org/abs/2303.08774

Panickssery, A., Bowman, S. R., & Feng, S. (2024). Llm evaluators recognize and favor their own generations. https://arxiv.org/abs/2404.13076

Peng, B., Li, C., He, P., Galley, M., & Gao, J. (2023). Instruction tuning with gpt-4.

Pillutla, K., Swayamdipta, S., Zellers, R., Thickstun, J., Welleck, S., Choi, Y., & Harchaoui, Z. (2021). Mauve: Measuring the gap between neural text and human text using divergence frontiers. https://arxiv.org/abs/2102.01454

Shumailov, I., Shumaylov, Z., Zhao, Y., et al. (2024). Ai models collapse when trained on recursively generated data. *Nature*, *631*, 755–759. https://doi.org/10.1038/s41586-024-07566-y

Singh, S., Vargus, F., Dsouza, D., Karlsson, B. F., Mahendiran, A., Ko, W.-Y., Shandilya, H., Patel, J., Mataciunas, D., OMahony, L., Zhang, M., Hettiarachchi, R., Wilson, J., Machado, M., Moura, L. S., Krzemiński, D., Fadaei, H., Ergün, I., Okoh, I., ... Hooker, S. (2024). Aya dataset: An open-access collection for multilingual instruction tuning. https://arxiv.org/abs/2402.06619

Taori, R., Gulrajani, I., Zhang, T., Dubois, Y., Li, X., Guestrin, C., Liang, P., & Hashimoto, T. B. (2023). Stanford alpaca: An instruction-following llama model.

Tian, Y., Peng, B., Song, L., Jin, L., Yu, D., Mi, H., & Yu, D. (2024). Toward self-improvement of llms via imagination, searching, and criticizing. https://arxiv.org/abs/2404.12253

Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., & Hobbhahn, M. (2024). Will we run out of data? limits of llm scaling based on human-generated data. https://arxiv.org/abs/2211.04325

Wang, H., Wang, S., Yan, Y., Wang, X., Yang, Z., Xu, Y., Liu, Z., Yang, L., Ding, N., Han, X., Liu, Z., & Sun, M. (2024). Ultralink: An open-source knowledge-enhanced multilingual supervised fine-tuning dataset. https://arxiv.org/abs/2402.04588

Xu, C., Sun, Q., Zheng, K., Geng, X., Zhao, P., Feng, J., Tao, C., & Jiang, D. (2023). Wizardlm: Empowering large language models to follow complex instructions. *ArXiv*, *abs/2304.12244*. https://api.semanticscholar.org/CorpusID:258298159

Xu, X., Li, M., Tao, C., Shen, T., Cheng, R., Li, J., Xu, C., Tao, D., & Zhou, T. (2024). A survey on knowledge distillation of large language models. https://arxiv.org/abs/2402.13116

Yang, A., Yang, B., Hui, B., Zheng, B., Yu, B., Zhou, C., Li, C., Li, C., Liu, D., Huang, F., Dong, G., Wei, H., Lin, H., Tang, J., Wang, J., Yang, J., Tu, J., Zhang, J., Ma, J., ... Fan, Z. (2024). Qwen2 technical report. https://arxiv.org/abs/2407.10671