

# Middo: Model-Informed Dynamic Data Optimization for Enhanced LLM Fine-Tuning via Closed-Loop Learning

Zinan Tang<sup>1,2</sup>, Xin Gao<sup>1,2</sup>, Qizhi Pei<sup>1,2</sup>, Zhuoshi Pan<sup>1,2</sup>  
Mengzhang Cai<sup>1,2</sup>, Jiang Wu<sup>1,2</sup>, Conghui He<sup>1,2\*</sup>, Lijun Wu<sup>1,2\*</sup>

<sup>1</sup>OpenDataLab <sup>2</sup>Shanghai Artificial Intelligence Laboratory  
tangzinan@pjlab.org.cn {heconghui,wulijun}@pjlab.org.cn

🤗 <https://huggingface.co/collections/Word2Li/Middo>

🐙 <https://github.com/Word2VecT/Middo>

## Abstract

Supervised Fine-Tuning (SFT) Large Language Models (LLMs) fundamentally rely on high-quality training data. While data selection and data synthesis are two common strategies to improve data quality, existing approaches often face limitations in static dataset curation that fail to adapt to evolving model capabilities. In this paper, we introduce **Middo**, a self-evolving **Model-informed dynamic data optimization** framework that uses model-aware data selection and context-preserving data refinement. Unlike conventional one-off filtering/synthesis methods, our framework establishes a closed-loop optimization system: (1) A self-referential diagnostic module proactively identifies suboptimal samples through tri-axial model signals—*loss patterns (complexity)*, *embedding cluster dynamics (diversity)*, and *self-alignment scores (quality)*; (2) An adaptive optimization engine then transforms suboptimal samples into pedagogically valuable training points while preserving semantic integrity; (3) This optimization process continuously evolves with the model’s capability through dynamic learning principles. Experiments on multiple benchmarks demonstrate that our Middo consistently enhances the quality of seed data and boosts LLMs’ performance, improving accuracy by 7.15% on average while maintaining the original dataset scale. This work establishes a new paradigm for sustainable LLM training through dynamic human-AI co-evolution of data and models.

## 1 Introduction

Large Language Models (LLMs) have revolutionized artificial intelligence by achieving state-of-the-art performance across diverse domains, from natural language understanding (Zhou et al., 2023c; Hendrycks et al., 2021b) to mathematical reasoning (Cobbe et al., 2021; Hendrycks et al., 2021a)

\* Corresponding authors.

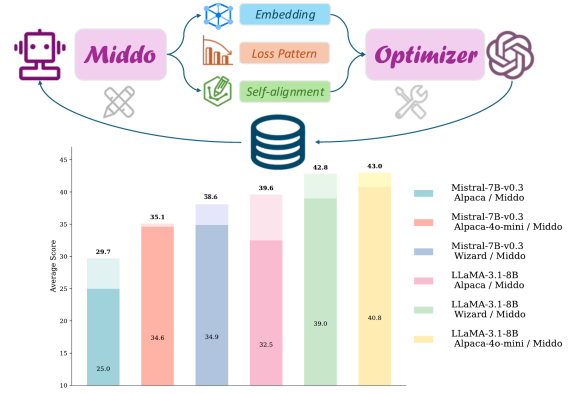


Figure 1: Comparison of different dataset and different models before and after Middo optimization.

and code generation (Chen et al., 2021; Austin et al., 2021). This success is largely attributed to Supervised Fine-Tuning (SFT), where models undergo rigorous training on high-quality, human-aligned datasets to ensure outputs closely match human expectations. Crucially, the quality of these datasets directly dictates the model’s ultimate capabilities: noisy or suboptimal training data can lead to degraded performance, while meticulously curated data unlocks advanced reasoning, generalization, and robustness. As LLMs scale, the adage “garbage in, garbage out” becomes increasingly important—highlighting the urgent need for systematic methods to optimize training data quality.

Existing approaches primarily fall into two categories to improve data quality: data selection (Cao et al., 2024; Zhou et al., 2023b; Li et al., 2024d; Jia et al., 2024; Zhou et al., 2025; Li et al., 2024e,a) and data synthesis (Dai et al., 2025; Wang et al., 2023a; Mukherjee et al., 2023; Xu et al., 2025; Liu et al., 2024a; Gao et al., 2025). Data selection methods filter raw datasets using heuristic rules (e.g., length filters) (Zhao et al., 2024a) or statistical metrics like perplexity (PPL) (Liu et al., 2024a) and Instruction-Following Difficulty (IFD) (Li et al., 2024c) to retain “high-quality” samples. Con-

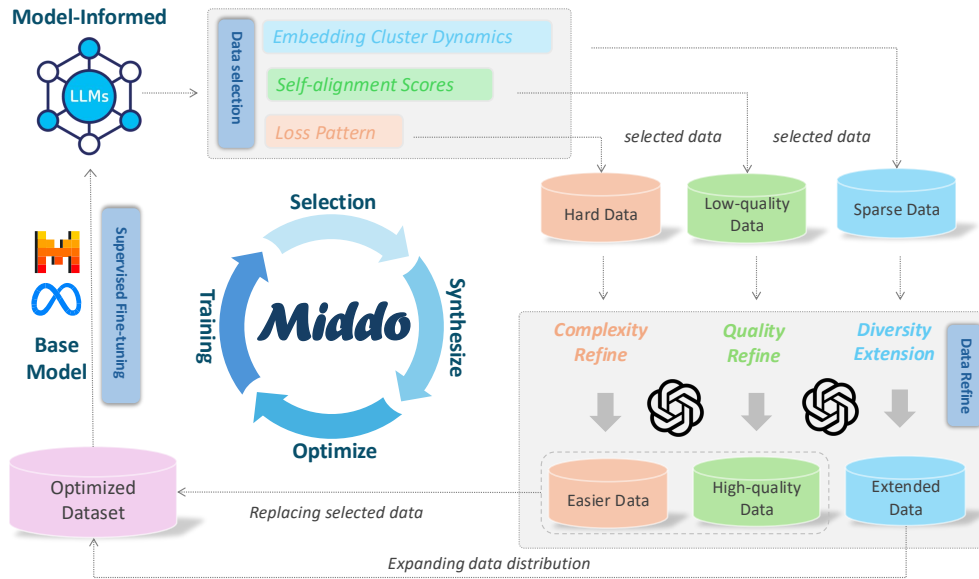


Figure 2: The Middo pipeline: a closed-loop, iterative dynamic optimization framework for LLM fine-tuning. It comprises three core modules that leverage model feedback: *Loss Patterns* identify overly complex samples, which are then simplified; *Self-alignment Scores* evaluate data quality, transforming low-quality samples into high-quality ones; and *Embedding Cluster Dynamics* detect sparse data points and expand the data distribution through targeted augmentation. Middo ensure the training set continually evolves to better align with the model’s capabilities.

versely, data synthesis leverages advanced LLMs (e.g., GPT-4 (Achiam et al., 2023)) to generate new training examples, often through prompting or distillation (Li et al., 2024f). While both strategies improve data quality, they suffer from critical limitations. Selection methods are typically static, applying fixed criteria that ignore the evolving needs of the model during training. Similarly, synthesis approaches often discard original data, wasting potentially valuable information, and risk generating distributionally narrow or redundant examples. These one-time data curation methods fail to adaptively refine data along with the model’s progress.

To overcome these limitations, we propose **Middo**, Model-informed Dynamic Data Optimization, a self-evolving framework that unifies model-aware data selection with context-preserving data refinement. Unlike static approaches, Middo establishes a closed-loop optimization system where data curation dynamically adapts to the model’s evolving capabilities. The framework operates through three core mechanisms: (1) A self-referential diagnostic module that proactively identifies suboptimal training samples using three model signals: *loss patterns* (to detect **complexity** mismatches between data and model proficiency), *embedding cluster dynamics* (to assess **diversity** gaps in the latent space), and *self-alignment scores*

(to evaluate data **quality** against the model’s own knowledge). (2) An adaptive optimization engine that transforms these suboptimal samples into pedagogically valuable training points. For example, overly complex samples may be simplified through stepwise decomposition, while low-diversity clusters are enriched with controlled extension—all while preserving the original data’s semantic intent. (3) A dynamic principle that iteratively updates the training dataset based on the model’s progress, ensuring that data difficulty and diversity scale with the model’s capabilities. By integrating these components, Middo not only maximizes the utility of existing data but also bridges the gap between static data curation and adaptive model training.

Experiments across multiple benchmarks demonstrate Middo’s effectiveness especially on low-quality datasets. Models trained with Middo optimized data achieve consistent performance gains over baselines, improving accuracy by 7.15% on average while maintaining the original dataset scale. Notably, Middo-trained models exhibit stronger abilities to address hard problems, solving more than three times the number of challenging test problems (e.g., MATH, GPQA) compared to models trained on static datasets. These results validate that sustainable LLM advancement requires co-evolving data and models—a paradigm shift

from today’s disjointed curation practices.

## 2 Related Work

### 2.1 Synthetic Data Generation

Synthetic data generation is a key technique for augmenting LLM fine-tuning. Early methods (Edunov et al., 2018; Wieting and Gimpel, 2018) introduce perturbation-based approaches to enhance data diversity, using character-level (Belinkov and Bisk, 2018) and word-level (Wei and Zou, 2019) modifications. These methods rely on fixed transformation rules, limiting adaptability.

LLMs have been leveraged for scalable data synthesis (Sudalairaj et al., 2024; Jung et al., 2024; Dai et al., 2025; Wang et al., 2023a; Mukherjee et al., 2023; Xu et al., 2025; Liu et al., 2024a; Li et al., 2025). Self-instruct methods (Wang et al., 2023b) generate instruction-response pairs, while Evol-Instruct (Xu et al., 2024) and Auto-Evol-Instruct (Zeng et al., 2024) refine data complexity iteratively. However, these methods remain static, failing to adapt as models improve. Recent approaches integrate model feedback into data generation (Khan et al., 2025; Maosongcao et al., 2025; Liang et al., 2025; Li et al., 2024f), incorporating student model signals for adaptive synthesis. LLM2LLM (Lee et al., 2024) is an iterative data augmentation strategy that enhances low-data fine-tuning by using a teacher LLM to generate synthetic training data from incorrect student LLM predictions and I-SHEEP (Park et al., 2024) uses an iterative self-enhancement paradigm.

### 2.2 Data Selection

Data selection is crucial for LLM fine-tuning, as high-quality and informative data directly impacts model performance (Zhou et al., 2023a; Xu et al., 2023). Early heuristic-based methods rely on surface-level statistics like item frequency (Raffel et al., 2020) and repetition count (Laurençon et al., 2022), but they also lack adaptability to model evolution.

Recent work explores LLM-driven data selection, optimizing for quality, diversity, and complexity (Cao et al., 2024; Zhou et al., 2023b; Li et al., 2024d; Jia et al., 2024; Zhou et al., 2025; Li et al., 2024e,a; Du et al., 2023; Kung et al., 2023). The IFD metric (Li et al., 2024c) enables models to self-select training instances by comparing loss with and without the instruction, while other methods (Yu et al., 2024; Colombo et al., 2024;

Lu et al., 2024; Zhao et al., 2024b) use LLM self-assessment for efficiency. Further advancements integrate LLM-based evaluation mechanisms. AlpGasus (Chen et al., 2024) and LIFT (Xu et al., 2023) use structured prompts for data assessment, while DEITA (Liu et al., 2024b) introduces a multi-dimensional scoring system based on complexity and quality.

## 3 Methodology

An overview of our Middo is shown in Figure 2. We first introduce the overall pipeline of Middo in Section 3.1, then elaborate on the three core components: complexity optimization (Section 3.2), diversity optimization (Section 3.3), and quality optimization (Section 3.4).

### 3.1 Middo Pipeline

As depicted in Figure 2, our Middo framework establishes an iterative data-model co-evolution loop driven by tri-axial signal analysis, along with three interconnected data optimization mechanisms, each targeting distinct dimensions of training sample selection: (1) *Loss patterns*, to identify samples with mismatched complexity (overly challenging) relative to the current model’s capability through loss trajectory analysis. (2) *Embedding cluster dynamics*, to detect coverage gaps in the semantic space, ensuring balanced conceptual representation. (3) *Self-alignment scores*, for quality filtering to leverage the model’s self-evaluation capacity to flag low-confidence or inconsistent responses through automated alignment scoring.

At each iteration, these parallel signal analyzers jointly select suboptimal samples, which are then regenerated through context-aware synthesis—preserving original semantic intent while enhancing pedagogical value. The refined dataset immediately feeds back into model training, creating a dynamic feedback loop where improved model capabilities inform subsequent optimization cycles. Notably, the optimized dataset remains similar in data size, without extending large data synthesis, leading to an efficient data optimization. This self-referential mechanism ensures continuous alignment between data characteristics and model evolution. The following sections systematically elaborate on the implementation of each signal-specific optimization module and their synergistic integration.

### 3.2 Loss Patterns: Complexity Optimization

**Complexity Selection.** Complexity reflects the “difficulty” or “compositionality” of data. A good dataset usually requires a smooth complexity distribution of data for training (Havrilla et al., 2024; Shen et al., 2024). Therefore, we introduce *Loss Patterns*, which targets overly challenging samples by modifying them to maintain a balanced and learnable training set (Zhao et al., 2024b). During fine-tuning, the loss for a sample  $(X_i, Y_i)$  is computed as the likelihood of predicting successive tokens given the instruction  $X_i$  and its context. We denote the loss before and after training as  $\mathcal{L}_{\text{pre}}(X_i, Y_i)$  and  $\mathcal{L}_{\text{post}}(X_i, Y_i)$ , respectively.

Intuitively, we consider both the loss before and after training to select the complex data. Specifically, we classify samples based on their loss evolution: samples with both low  $\mathcal{L}_{\text{pre}}$  and  $\mathcal{L}_{\text{post}}$  are considered easy, while those with high values in both metrics remain difficult, indicating excessive complexity. A sample is included in the complex subset  $\mathcal{D}^{\text{hard}}$  if its  $\mathcal{L}_{\text{pre}}$  and  $\mathcal{L}_{\text{post}}$  both exceed the thresholds  $\tau_{\text{pre}}$  and  $\tau_{\text{post}}$ , respectively. For adaptive refinement, the thresholds are dynamically computed. See Appendix B.3 for details on the dynamic threshold settings used throughout the paper.

**Complexity Optimization.** For complex data optimization, instead of discarding difficult samples, we transform them into simpler, more manageable forms. Specifically, we replace samples in  $\mathcal{D}^{\text{hard}}$  with their simplified counterparts,  $\mathcal{D}^{\text{hard}'}$ . This is achieved by an automatic process in which a LLM analyzes and summarizes the complex instructions (Zeng et al., 2024), then simplifies them step by step while preserving the core educational content. An example is shown in Appendix Figure 11. This iterative transformation process updates the dataset by replacing overly complex samples with refined versions that offer more effective training samples. As training continues, this adaptive approach ensures a continuous alignment between data complexity and model capability.

### 3.3 Embedding Cluster Dynamics: Diversity Optimization

**Diversity Selection.** Diversity is crucial for ensuring broad concept coverage and a uniform data distribution. *Embedding Cluster Dynamics* identifies sparse data points that signal underrepresented regions in the dataset. We extract sentence embeddings from the last hidden layer  $\mathcal{H}^{(L)}$  of the model

trained in the previous iteration, using average pooling, then compute the cosine similarity between each data point and find the  $k$ -nearest neighbors  $\mathcal{N}_k(X_i)$  for each data  $X_i$ . A lower average cosine similarity among these neighbors  $\mathcal{N}_k(X_i)$  indicates the data is positioned in a sparser region. Thus, the data points whose average cosine similarity score (diversity score) is below a threshold are selected for optimization.

**Diversity Optimization.** To enhance diversity-balanced distribution, we augment the sparse subset  $\mathcal{D}^{\text{sparse}}$  by incorporating examples from their corresponding  $\mathcal{N}_k(X_i)$  as demonstrations to generate new samples. This process generates an expanded set  $\mathcal{D}^{\text{sparse}'}$ , which is then integrated back into the dataset. An instance can be found in Appendix Figure 13. This structured augmentation strategy ensures that the data distribution becomes both broader and more balanced, ultimately improving the model’s generalization.

### 3.4 Self-alignment Scores: Quality Optimization

**Quality Selection.** High-quality data is essential for fine-tuning, as poor-quality samples can degrade performance (Zhou et al., 2023a). To reduce manual annotation costs, many approaches use the LLM-as-a-Judge paradigm (Chen et al., 2024; Xu et al., 2023). To achieve this, instead of relying on an external judge, we leverage the fine-tuned model itself to assess data quality via *Self-alignment Scores*, effectively incorporating the model’s own feedback. Specifically, for each instruction-response pair  $(X_i, Y_i)$  in  $\mathcal{D}$ , the model predicts scores  $\mathcal{S}\pi^{\text{ins.}}(X_i)$  for instruction and  $\mathcal{S}\pi^{\text{res.}}(X_i, Y_i)$  for instruction-response pair based on three key metrics  $\pi$  from AlignBench (Liu et al., 2024c): *Clarity*, *Completeness*, and *Factuality*. The final quality score  $\mathcal{S}(X_i, Y_i)$  is obtained by averaging these scores. These samples with scores below a similar dynamic threshold are identified as low-quality, forming the seed dataset  $\mathcal{D}^{\text{low}}$ .

**Quality Optimization.** To refine  $\mathcal{D}^{\text{low}}$ , we use LLMs to automatically analyze and improve these samples via tailored evolution strategies (prompt templates and examples are provided in the Appendix Figure 12). This process converts low-quality samples into higher-quality versions, denoted as  $\mathcal{D}^{\text{low}'}$ . The dataset is then updated by replacing the original low-quality samples with  $\mathcal{D}^{\text{low}'}$ ,

maintaining the dataset size while progressively enhancing its overall quality.

In each iteration, after the three data selection and optimization processes described above, the optimized dataset is then fed back for the next round of model training.

## 4 Experiment

### 4.1 Settings

**Data Optimization Configurations.** We conduct optimization on the Alpaca (Taori et al., 2023) and WizardLM (Xu et al., 2024) datasets. For a fair comparison, we also include a rewritten version of Alpaca, where responses are generated by GPT-4o-mini, in our optimization process. Each dataset undergoes three iterations of optimization. Demonstrating that our method does not require a powerful external model, we synthesize data using DataDremer (Patel et al., 2024) with GPT-4o-mini, setting both temperature and top\_p to 1.0 to ensure diversity. A detailed analysis of the computational cost is provided in Appendix A, and the effects of the number of neighbors and iteration counts are discussed in Appendix B.

**Training and Evaluation Settings.** We fine-tune LLaMA-3.1-8B (Dubey et al., 2024) and Mistral-7B-v0.3 (Jiang et al., 2023) using LLaMA-Factory (Zheng et al., 2024) with the specific hyperparameters detailed in Appendix C.5. For each iteration of Middo’s optimization, the base model is fine-tuned for one epoch on the dataset optimized in that specific iteration to mitigate the risk of overfitting to the data (Lee et al., 2024; Liang et al., 2025). Evaluation is conducted using OpenCompass (Contributors, 2023), with vLLM (Kwon et al., 2023) for acceleration. To validate the effectiveness and generalization capabilities of our approach, we assess model capabilities in general knowledge using IFEval (Zhou et al., 2023c) and MMLU (Hendrycks et al., 2021b); mathematical problem-solving on GSM8K (Cobbe et al., 2021) and MATH (Hendrycks et al., 2021a); code generation on HumanEval (Chen et al., 2021) and MBPP (Austin et al., 2021); and commonsense reasoning on Hellaswag (Zellers et al., 2019) and GPQA (Rein et al., 2024).

### 4.2 Main Results

The evaluation results on all benchmarks over various data iterations and models are presented in

Table 1. We can see that Middo consistently enhances model performance across all benchmarks, achieving an average accuracy increase of 7.15% over three iterations on the Alpaca dataset based on LLaMA-3.1-8B, all while preserving the original data scale. Moreover, when extending our experiments to Mistral-7B-v0.3, we observed an average improvement of 4.75%, further underscoring the robustness and adaptability of our framework across different model architectures.

**Dynamic Iterative Improvement.** On the Alpaca dataset, the average score increased progressively with each iteration. Across the MMLU, GSM8K, MATH, and MBPP benchmarks, we observed consistent, step-by-step improvements over multiple iterations. This showcases the versatility of our approach, which excels in general capabilities, mathematics, and coding. Notably, accuracy on GSM8K improved by 15.55%, and Hellaswag saw an 11.11% increase when evaluated on the LLaMA-3.1-8B model. For Mistral-7B-v0.3, we observed an 11.07% improvement on MMLU, a 12.59% increase on GSM8K, and a 10.6% gain on GPQA. These results underscore the effectiveness of our method in driving performance gains and highlight the cumulative benefit of our iterative optimization process.

**Further Validation on 4o-mini Rewritten Data.** Steady improvements observed on the 4o-mini rewritten Alpaca dataset—averaging a 2.2% increase overall, with MMLU showing an impressive 11.87% boost—demonstrate that these gains are not merely a result of using 4o-mini data. This illustrates that our framework intrinsically enhances data quality and model performance. Importantly, we achieve these improvements without resorting to stronger variants such as GPT-4o (Hurst et al., 2024), reinforcing the robustness and general applicability of our method.

**Initial Dataset Quality.** Our experiments reveal that higher-quality datasets require fewer modifications to reach optimal performance. On LLaMA-3.1-8B, for instance, while the Alpaca dataset achieves peak performance at third iteration, the 4o-mini rewritten Alpaca required only two iterations, and the Wizard dataset reaches its best performance in just one round.

**Comparison with Other Works.** We compare Middo with both data selection (Alpaca-clean (Ruebsamen, 2023), Superfiltering (Li et al.,

| Setting                     |       | General |        | Math  |       | Code      |       | Reasoning |       | Average      |
|-----------------------------|-------|---------|--------|-------|-------|-----------|-------|-----------|-------|--------------|
|                             |       | MMLU    | IFEval | GSM8K | MATH  | HumanEval | MBPP  | Hellaswag | GPQA  |              |
| Base Model: LLaMA-3.1-8B    |       |         |        |       |       |           |       |           |       |              |
| Alpaca                      | init  | 47.46   | 41.09  | 35.63 | 4.96  | 39.63     | 37.40 | 48.11     | 5.56  | 32.48        |
|                             | iter1 | 50.13   | 45.77  | 43.67 | 10.62 | 40.24     | 39.20 | 56.37     | 13.64 | 37.45        |
|                             | iter2 | 41.82   | 44.63  | 50.11 | 12.40 | 39.63     | 41.40 | 59.22     | 18.18 | <u>38.42</u> |
|                             | iter3 | 51.32   | 43.20  | 51.18 | 12.92 | 39.63     | 41.80 | 58.78     | 16.67 | <b>39.63</b> |
| Alpaca 4o-mini              | init  | 32.82   | 44.04  | 57.09 | 17.78 | 51.22     | 45.20 | 53.70     | 24.24 | 40.76        |
|                             | iter1 | 41.09   | 43.47  | 54.21 | 17.34 | 51.22     | 46.00 | 59.11     | 21.72 | 41.77        |
|                             | iter2 | 44.69   | 47.96  | 57.62 | 18.50 | 52.44     | 45.40 | 57.37     | 19.70 | <b>42.96</b> |
|                             | iter3 | 38.58   | 48.11  | 58.68 | 18.30 | 46.95     | 46.80 | 52.37     | 28.79 | <u>42.32</u> |
| Wizard                      | init  | 46.12   | 46.14  | 53.30 | 12.72 | 40.24     | 48.00 | 53.05     | 12.12 | 38.96        |
|                             | iter1 | 48.39   | 50.11  | 54.44 | 13.80 | 46.95     | 45.00 | 63.54     | 20.20 | <b>42.80</b> |
|                             | iter2 | 48.86   | 49.48  | 55.12 | 13.90 | 48.78     | 45.20 | 58.63     | 18.18 | <u>42.29</u> |
|                             | iter3 | 47.18   | 50.79  | 54.51 | 11.70 | 43.29     | 45.40 | 62.97     | 20.20 | 42.01        |
| Base Model: Mistral-7B-v0.3 |       |         |        |       |       |           |       |           |       |              |
| Alpaca                      | init  | 27.66   | 43.22  | 22.21 | 3.88  | 29.27     | 28.80 | 44.17     | 0.51  | 24.97        |
|                             | iter1 | 31.31   | 45.62  | 29.57 | 5.82  | 30.49     | 33.80 | 42.73     | 14.65 | 29.25        |
|                             | iter2 | 26.87   | 49.46  | 31.69 | 6.84  | 31.71     | 31.00 | 53.95     | 5.56  | <u>29.64</u> |
|                             | iter3 | 38.73   | 44.01  | 34.80 | 6.64  | 26.22     | 31.40 | 44.86     | 11.11 | <b>29.72</b> |
| Alpaca 4o-mini              | init  | 31.56   | 43.14  | 44.88 | 9.64  | 42.07     | 37.80 | 46.25     | 21.21 | 34.56        |
|                             | iter1 | 31.33   | 47.93  | 45.19 | 8.72  | 37.20     | 41.32 | 41.32     | 19.70 | 34.09        |
|                             | iter2 | 28.83   | 47.92  | 48.90 | 11.34 | 35.37     | 38.40 | 42.63     | 27.27 | <b>35.08</b> |
|                             | iter3 | 28.96   | 50.78  | 48.60 | 10.10 | 32.32     | 39.00 | 32.95     | 20.20 | 32.86        |
| Wizard                      | init  | 40.71   | 50.95  | 44.96 | 8.10  | 35.98     | 35.60 | 53.98     | 9.09  | 34.92        |
|                             | iter1 | 41.39   | 51.18  | 44.43 | 9.44  | 37.80     | 38.60 | 59.01     | 17.17 | 37.38        |
|                             | iter2 | 33.87   | 51.71  | 47.08 | 9.26  | 39.02     | 38.40 | 66.18     | 19.70 | <u>38.15</u> |
|                             | iter3 | 43.26   | 49.80  | 41.09 | 10.02 | 41.46     | 34.60 | 66.02     | 22.22 | <b>38.56</b> |

Table 1: Performance comparison on different benchmarks using *LLaMA-3.1-8B* and *Mistral-7B-v0.3* as base models. We use Alpaca, Alpaca-4o-mini, and Wizard as the optimization datasets for Mido. The *init* means training on the original dataset, while *iter* means training on the optimized dataset. Both *init* and *iter* settings are trained for one epoch. The best performance on average is highlighted in **bold** and the second best is underlined.

| Method                                     | Size  | General |        | Math  |       | Code      |       | Reasoning |       | Average      |
|--------------------------------------------|-------|---------|--------|-------|-------|-----------|-------|-----------|-------|--------------|
|                                            |       | MMLU    | IFEval | GSM8K | MATH  | HumanEval | MBPP  | Hellaswag | GPQA  |              |
| Alpaca                                     | 52.0k | 47.46   | 41.09  | 35.63 | 4.96  | 39.63     | 37.40 | 48.11     | 5.56  | 32.48        |
| <i>Data Selection</i>                      |       |         |        |       |       |           |       |           |       |              |
| Alpaca-clean                               | 51.7k | 47.21   | 43.92  | 43.90 | 4.20  | 29.27     | 43.40 | 60.17     | 5.56  | 34.70        |
| Superfiltering                             | 7.8k  | 39.96   | 37.80  | 44.50 | 5.38  | 40.85     | 44.00 | 42.38     | 27.27 | 35.27        |
| Superfiltering GPT4                        | 7.8k  | 37.71   | 34.35  | 53.68 | 11.00 | 9.15      | 45.60 | 57.81     | 2.53  | 31.48        |
| Long                                       | 1.0k  | 25.51   | 14.75  | 56.33 | 16.56 | 13.41     | 45.60 | 25.83     | 0.00* | 24.75        |
| AlpaGasus                                  | 9.2k  | 33.98   | 48.82  | 43.82 | 6.06  | 35.98     | 42.40 | 44.50     | 18.18 | 34.22        |
| <i>Data Augmentation</i>                   |       |         |        |       |       |           |       |           |       |              |
| I-SHEEP                                    | 8.4k  | 23.61   | 29.61  | 43.14 | 8.28  | 32.32     | 32.60 | 41.83     | 0.00* | 26.42        |
| Alpaca-GPT4                                | 52.0k | 51.94   | 38.68  | 50.87 | 10.28 | 17.07     | 43.60 | 63.02     | 0.51  | 34.50        |
| WizardLM                                   | 70.0k | 46.12   | 46.14  | 53.30 | 12.72 | 40.24     | 48.00 | 53.05     | 12.12 | 38.96        |
| <i>Middo Alpaca</i>                        | 57.6k | 51.32   | 43.20  | 51.18 | 12.92 | 39.63     | 41.80 | 58.78     | 16.67 | 39.63        |
| <i>MiddOnly<sup>†</sup> Alpaca</i>         | 8.8k  | 43.47   | 40.78  | 65.20 | 15.58 | 51.83     | 47.60 | 58.65     | 17.68 | 42.60        |
| <i>Middo Alpaca-4o-mini</i>                | 63.1k | 44.69   | 47.96  | 57.62 | 18.50 | 52.44     | 45.40 | 57.37     | 19.70 | <b>42.96</b> |
| <i>MiddOnly<sup>†</sup> Alpaca-4o-mini</i> | 24.9k | 41.50   | 45.66  | 60.80 | 20.06 | 46.34     | 48.00 | 55.01     | 24.75 | <u>42.77</u> |

Table 2: Results of Mido compared to other baseline methods. The best and second best results are highlighted in **bold** and underlined, respectively. Our method outperforms all baselines in the average score. \*Note that 0.00 indicates that the method did not solve any examples. <sup>†</sup> Denotes training solely on Mido-generated data.

2024b), Long (Zhao et al., 2024a), AlpaGauss (Chen et al., 2024)) and data augmentation (Alpaca-GPT4 (Peng et al., 2023), I-SHEEP (Liang et al., 2025), WizardLM (Xu et al., 2024)) methods on the Alpaca dataset.

We use the optimal dataset obtained through Middo from Alpaca for comparison with other baselines. Additionally, to ensure a relatively fair comparison with data selection methods, we include a dataset that only uses the optimized data without incorporating any unoptimized samples, referred to as *MiddOnly*, to isolate the effect of the optimization process and make a direct comparison with data selection approaches.

Results in Table 2 show our method achieves the highest average score of 42.96, outperforming all other approaches. Notably, even when using only the optimized subset *MiddOnly Alpaca*, our method delivers a robust average score of 42.60. This demonstrates that iterative improvement is not primarily driven by data size, but rather by the effectiveness of our dynamic data selection and optimization process in identifying and generating data with high learning value for models.

## 5 Analysis

### 5.1 Ablation Studies

To assess the effectiveness of Middo and the contribution of each optimization pipeline, we conduct ablation experiments with the LLaMA-3.1-8B model on the Alpaca dataset. Specifically, we analyze the following ablations: **(a) w/o loss**: removes *Loss Patterns*. **(b) w/o neighbor**: excludes *Embedding Cluster Dynamics*. **(c) w/o score**: removes *Self-alignment Scores*.

| iteration    | Ablations    | IFEval | MATH  | HumanEval | Hellaswag | Average      |
|--------------|--------------|--------|-------|-----------|-----------|--------------|
| <i>iter1</i> | w            | 45.77  | 10.62 | 40.24     | 56.37     | <b>38.25</b> |
|              | w/o loss     | 42.49  | 10.11 | 39.02     | 59.53     | 37.79        |
|              | w/o neighbor | 39.01  | 10.82 | 42.07     | 57.86     | 37.45        |
|              | w/o score    | 43.48  | 10.20 | 36.59     | 48.40     | 34.67        |
| <i>iter2</i> | w            | 44.63  | 12.40 | 39.63     | 59.22     | <b>38.97</b> |
|              | w/o loss     | 42.28  | 9.92  | 42.68     | 58.21     | 38.27        |
|              | w/o neighbor | 46.75  | 10.26 | 34.76     | 46.66     | 34.61        |
|              | w/o score    | 44.18  | 11.76 | 39.02     | 51.38     | 36.58        |
| <i>iter3</i> | w            | 44.24  | 12.92 | 39.63     | 59.25     | <b>39.01</b> |
|              | w/o loss     | 43.18  | 12.42 | 36.59     | 55.30     | 36.87        |
|              | w/o neighbor | 40.12  | 12.46 | 34.15     | 56.83     | 35.89        |
|              | w/o score    | 45.17  | 7.92  | 40.85     | 54.67     | 37.15        |

Table 3: Ablation study on the development set. We report the performance of the model with different ablations. The ablations include removing the *loss patterns*, *embedding cluster dynamics* and *self-alignment scores* separately. The best performance is highlighted in **bold**.

The ablation results in Table 3 consistently show

that removing any part of the framework leads to a decline in performance across multiple iterations, reinforcing that each component plays a significant role in the overall performance. This trend holds across the second (*iter2*) and third (*iter3*) iterations, where the removal of any pipeline consistently results in suboptimal performance, further highlighting the importance of balancing complexity, diversity, and quality in the optimization process. These findings underscore the necessity of the full framework for achieving optimal results.

### 5.2 Effect of Selected Data Scale

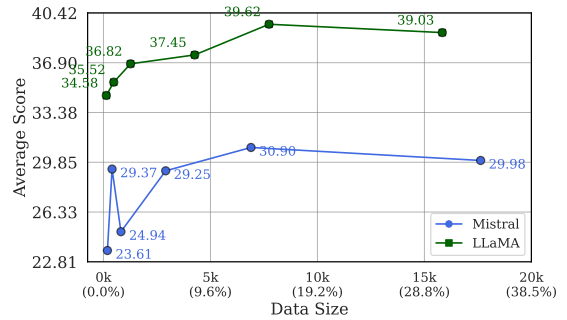


Figure 3: Performance comparison of Middo on the Alpaca dataset with varying refined data sizes. The x-axis represents the number and percentage of data selected for refinement, while the y-axis shows the average accuracy across three iterations. To ensure fairness, we guarantee that the data after refinement is the same.

We investigate the impact of the different scales of the selected and optimized data in this section by varying the thresholds for data selection. Results are illustrated in Figure 3. We observe that increasing the size of the refined data initially leads to an upward trend in performance; however, once the refined data exceeds a certain threshold, performance begins to decline. To maintain the potential for further iterative improvement, we set the refined data size at a moderate level that optimally balances the cost and benefit of the optimization process. In the first iteration, each component selects approximately 5% of the data for refinement. By controlling the parameter  $m$ , the amount of data refined can adaptively change as the model’s capability increases. Detailed data sizes selected in each iteration are provided in Appendix E.

### 5.3 Data Analysis

For a deeper understanding of how Middo transforms the dataset, we provide an analysis of its impact on data complexity, diversity, and quality.

**Complexity.** To quantify how Middo modulates dataset complexity, we analyze the loss distribution evolution through optimization cycles. As shown in Figure 4, the original dataset exhibits a long-tailed distribution with extreme loss values up to 12.99. After applying Middo, the maximum loss decreases by 71.05% to 3.76, indicating successful mitigation of overly complex samples and the distribution mode shifts leftward, suggesting better alignment between data complexity and model capability. This transformation demonstrates our framework’s ability to adaptively prune pathological samples while preserving pedagogically valuable challenges.

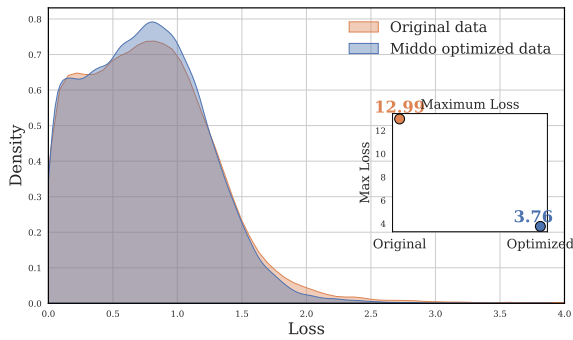


Figure 4: Loss distribution comparison before and after applying Middo. The density curve reflects the relative frequency of data points within specific loss intervals. The inset subfigure highlights the maximum loss reduction from 12.99 to 3.76.

**Diversity.** To analyze the diversity of the dataset after applying Middo, we visualize the data distribution using t-SNE (Van der Maaten and Hinton, 2008). Figure 5 reveals how the augmented data points are distributed relative to the original data. Notably, most of the augmented samples are located at the peripheries of the clusters, effectively filling in the sparsely populated regions. This distribution indicates that Middo is not merely adding redundant data but is instead enhancing the overall coverage of the latent space. By strategically augmenting the dataset at the cluster edges, Middo improves the diversity and ensures a more uniform distribution of data points, ultimately contributing to better model generalization.

**Quality.** The self-alignment score trajectories across different iterations are presented in Figure 6. The observed trend indicates a gradual increase in the average score as the iterations progress. This improvement signifies that the quality of the data

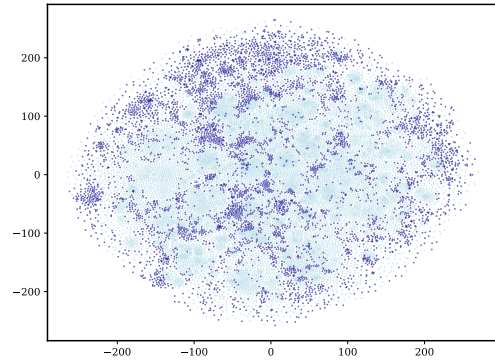


Figure 5: t-SNE visualization of the Alpaca dataset before and after applying Middo. The original dataset is shown in light blue, while the augmented data is in dark blue. The dark blue points tend to occupy the sparsely populated regions of the light blue point distribution.

is becoming more closely aligned with the model’s evolving capabilities. Through the adversarial self-play mechanisms and iterative quality refinement, the model is able to assess and enhance the quality of both the instructions and responses within the dataset. As the self-alignment scores increase, it reflects that the refined data is not only more accurate but also more consistent with the model’s internal standards and expectations. This detailed evolution of the self-alignment scores provides critical insights into the dynamic process of dataset optimization, confirming that our approach effectively transforms low-quality samples into high-quality learning material over successive iterations.

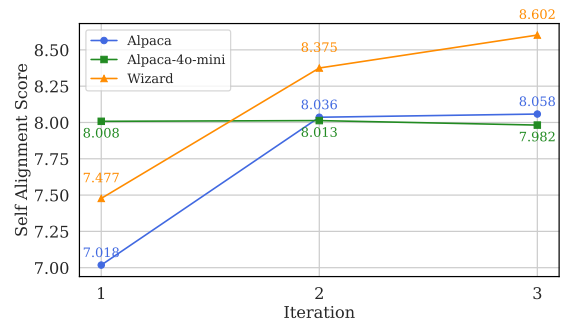


Figure 6: Self-alignment score evolution across iterations. The x-axis represents the number of iterations, while the y-axis shows the average self-alignment score.

## 6 Conclusion

In this paper, we present Middo, a model-informed dynamic data optimization framework that trans-

forms LLM fine-tuning via closed-loop learning. Unlike traditional static methods, Middo establishes a self-evolving system that continuously adapts to the model’s evolving capabilities. It employs three core mechanisms: complexity optimization refines overly complex samples using *loss patterns*, ensuring the training data remains appropriately challenging; diversity optimization enhances dataset diversity by analyzing *embedding cluster dynamics*; and quality optimization leverages *self-alignment scores* to evaluate and improve the quality of training samples. Experiments on multiple benchmarks demonstrate that Middo consistently boosts LLMs’ performance, achieving an average accuracy improvement of 7.15% while maintaining the original data scale on LLaMA-3.1-8B. Ablation studies confirm the effectiveness of each component, underscoring the importance of balancing complexity, diversity, and quality. Middo’s adaptability and model-awareness make it a powerful tool for sustainable LLM training. Moreover, our approach paves the way for future research in adaptive training that continuously optimizes learning efficiency.

## Limitations

Despite its promising results, Middo has several limitations: (1) Middo relies on the model being fine-tuned itself for identifying data quality and complexity. This means that the approach requires a sufficiently capable base model, and the performance may be limited if the base model is not strong enough to generate meaningful diagnostics for data refinement. (2) Middo does not currently utilize Reinforcement Learning, which could further enhance data refinement, especially for complex or subjective tasks. (3) The closed-loop optimization system may lead to higher computational costs as the dataset grows or updates become more frequent, presenting scalability challenges. (4) Middo may propagate biases present in the initial training data, limiting fairness and generalization if the base model is trained on biased data. These limitations highlight areas for future improvement, such as integrating RL, optimizing for scalability, and addressing data biases.

## Acknowledgments

This work is supported by Shanghai Artificial Intelligence Laboratory. Zinan Tang is an intern at Shanghai Artificial Intelligence Laboratory.

## References

- Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and Charles Sutton. 2021. [Program synthesis with large language models](#). *Preprint*, arXiv:2108.07732.
- Yonatan Belinkov and Yonatan Bisk. 2018. [Synthetic and natural noise both break neural machine translation](#). In *International Conference on Learning Representations*.
- Yihan Cao, Yanbin Kang, Chi Wang, and Lichao Sun. 2024. [Instruction mining: Instruction data selection for tuning large language models](#). In *First Conference on Language Modeling*.
- Lichang Chen, Shiyang Li, Jun Yan, Hai Wang, Kalpa Gunaratna, Vikas Yadav, Zheng Tang, Vijay Srinivasan, Tianyi Zhou, Heng Huang, and Hongxia Jin. 2024. [Alpagasus: Training a better alpaca with fewer data](#). In *The Twelfth International Conference on Learning Representations*.
- Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgen Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N. Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. 2021. [Evaluating large language models trained on code](#). *Preprint*, arXiv:2107.03374.
- Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. 2021. [Training verifiers to solve math word problems](#). *Preprint*, arXiv:2110.14168.
- Pierre Colombo, Telmo Pessoa Pires, Malik Boudiaf, Dominic Culver, Rui Melo, Caio Corro, Andre F. T. Martins, Fabrizio Esposito, Vera Lúcia Raposo, Sofia Morgado, and Michael Desa. 2024. [SaulLM-7b: A pioneering large language model for law](#). *Preprint*, arXiv:2403.03883.

- OpenCompass Contributors. 2023. Opencompass: A universal evaluation platform for foundation models. <https://github.com/open-compass/opencompass>.
- Haixing Dai, Zhengliang Liu, Wenxiong Liao, Xiaoke Huang, Yihan Cao, Zihao Wu, Lin Zhao, Shaochen Xu, Fang Zeng, Wei Liu, et al. 2025. Auggpt: Leveraging chatgpt for text data augmentation. *IEEE Transactions on Big Data*.
- Qianlong Du, Chengqing Zong, and Jiajun Zhang. 2023. **Mods: Model-oriented data selection for instruction tuning**. *Preprint*, arXiv:2311.15653.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.
- Sergey Edunov, Myle Ott, Michael Auli, and David Grangier. 2018. Understanding back-translation at scale. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 489–500.
- Xin Gao, Qizhi Pei, Zinan Tang, Yu Li, Honglin Lin, Jiang Wu, Lijun Wu, and Conghui He. 2025. **A strategic coordination framework of small LMs matches large LMs in data synthesis**. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11552–11570, Vienna, Austria. Association for Computational Linguistics.
- Alex Havrilla, Andrew Dai, Laura O’Mahony, Koen Oostermeijer, Vera Zisler, Alon Albalak, Fabrizio Milo, Sharath Chandra Raparthy, Kanishk Gandhi, Baber Abbasi, et al. 2024. Surveying the effects of quality, diversity, and complexity in synthetic data from large language models. *arXiv preprint arXiv:2412.02980*.
- Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021a. **Measuring massive multitask language understanding**. In *International Conference on Learning Representations*.
- Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. 2021b. **Measuring mathematical problem solving with the MATH dataset**. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.
- Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.
- Qi Jia, Siyu Ren, Ziheng Qin, Fuzhao Xue, Jinjie Ni, and Yang You. 2024. **Boosting llm via learning from data iteratively and selectively**. *Preprint*, arXiv:2412.17365.
- Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, L  lio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timoth  e Lacroix, and William El Sayed. 2023. **Mistral 7b**. *Preprint*, arXiv:2310.06825.
- Jaehun Jung, Peter West, Liwei Jiang, Faeze Brahman, Ximing Lu, Jillian Fisher, Taylor Sorensen, and Yejin Choi. 2024. Impossible distillation for paraphrasing and summarization: How to make high-quality lemonade out of small, low-quality model. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4439–4454.
- Zaid Khan, Elias Stengel-Eskin, Jaemin Cho, and Mohit Bansal. 2025. **Dataenvgym: Data generation agents in teacher environments with student feedback**. In *The Thirteenth International Conference on Learning Representations*.
- Po-Nien Kung, Fan Yin, Di Wu, Kai-Wei Chang, and Nanyun Peng. 2023. Active instruction tuning: Improving cross-task generalization by training on prompt sensitive tasks. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1813–1829.
- Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th symposium on operating systems principles*, pages 611–626.
- Hugo Lauren  on, Lucile Saulnier, Thomas Wang, Christopher Akiki, Albert Villanova del Moral, Teven Le Scao, Leandro Von Werra, Chenghao Mou, Eduardo Gonz  lez Ponferrada, Huu Nguyen, et al. 2022. The bigscience roots corpus: A 1.6 tb composite multilingual dataset. *Advances in Neural Information Processing Systems*, 35:31809–31826.
- Nicholas Lee, Thanakul Wattanawong, Sehoon Kim, Karttikeya Mangalam, Sheng Shen, Gopala Anumanchipalli, Michael Mahoney, Kurt Keutzer, and Amir Gholami. 2024. Llm2llm: Boosting llms with novel iterative data enhancement. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 6498–6526.
- Haoran Li, Qingxiu Dong, Zhengyang Tang, Chaojun Wang, Xingxing Zhang, Haoyang Huang, Shaohan Huang, Xiaolong Huang, Zeqiang Huang, Dongdong Zhang, Yuxian Gu, Xin Cheng, Xun Wang, Si-Qing Chen, Li Dong, Wei Lu, Zhifang Sui, Benyou Wang, Wai Lam, and Furu Wei. 2025. **Synthetic data (almost) from scratch: Generalized instruction tuning for language models**. *Transactions on Machine Learning Research*.

- Ming Li, Lichang Chen, Jiuhai Chen, Shwai He, Jixiang Gu, and Tianyi Zhou. 2024a. Selective reflection-tuning: Student-selected data recycling for llm instruction-tuning. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 16189–16211.
- Ming Li, Yong Zhang, Shwai He, Zhitao Li, Hongyu Zhao, Jianzong Wang, Ning Cheng, and Tianyi Zhou. 2024b. Superfiltering: Weak-to-strong data filtering for fast instruction-tuning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14255–14273.
- Ming Li, Yong Zhang, Zhitao Li, Jiuhai Chen, Lichang Chen, Ning Cheng, Jianzong Wang, Tianyi Zhou, and Jing Xiao. 2024c. From quantity to quality: Boosting llm performance with self-guided data selection for instruction tuning. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 7595–7628.
- Xian Li, Ping Yu, Chunting Zhou, Timo Schick, Omer Levy, Luke Zettlemoyer, Jason E Weston, and Mike Lewis. 2024d. [Self-alignment with instruction back-translation](#). In *The Twelfth International Conference on Learning Representations*.
- Yunshui Li, Binyuan Hui, Xiaobo Xia, Jiayi Yang, Min Yang, Lei Zhang, Shuzheng Si, Ling-Hao Chen, Junhao Liu, Tongliang Liu, et al. 2024e. One-shot learning as instruction data prospector for large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4586–4601.
- Zhuochun Li, Yuelyu Ji, Rui Meng, and Daqing He. 2024f. [Learning from committee: Reasoning distillation from a mixture of teachers with peer-review](#). Preprint, arXiv:2410.03663.
- Yiming Liang, Ge Zhang, Xingwei Qu, Tianyu Zheng, Xeron Du, Jiawei Guo, Zhenzhu Yang, Jiaheng Liu, Chenghua Lin, Lei Ma, Stephen Huang, and Jiajun Zhang. 2025. [I-SHEEP: Self-alignment of LLM from scratch through an iterative self-enhancement paradigm](#). In *Scaling Self-Improving Foundation Models without Human Supervision*.
- Ruibo Liu, Jerry Wei, Fangyu Liu, Chenglei Si, Yanzhe Zhang, Jinmeng Rao, Steven Zheng, Daiyi Peng, Diyi Yang, Denny Zhou, and Andrew M. Dai. 2024a. [Best practices and lessons learned on synthetic data](#). In *First Conference on Language Modeling*.
- Wei Liu, Weihao Zeng, Keqing He, Yong Jiang, and Junxian He. 2024b. [What makes good data for alignment? a comprehensive study of automatic data selection in instruction tuning](#). In *The Twelfth International Conference on Learning Representations*.
- Xiao Liu, Xuanyu Lei, Shengyuan Wang, Yue Huang, Andrew Feng, Bosi Wen, Jiale Cheng, Pei Ke, Yifan Xu, Weng Lam Tam, et al. 2024c. [Alignbench: Benchmarking chinese alignment of large language models](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11621–11640.
- Keming Lu, Hongyi Yuan, Zheng Yuan, Runji Lin, Junyang Lin, Chuanqi Tan, Chang Zhou, and Jingren Zhou. 2024. [#instag: Instruction tagging for analyzing supervised fine-tuning of large language models](#). In *The Twelfth International Conference on Learning Representations*.
- Maosongcao Maosongcao, Taolin Zhang, Mo Li, Chuyu Zhang, Yunxin Liu, Conghui He, Haodong Duan, Songyang Zhang, and Kai Chen. 2025. Condor: Enhance llm alignment with knowledge-driven data synthesis and refinement. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 22392–22412.
- Subhabrata Mukherjee, Arindam Mitra, Ganesh Jawahar, Sahaj Agarwal, Hamid Palangi, and Ahmed Awadallah. 2023. [Orca: Progressive learning from complex explanation traces of gpt-4](#). Preprint, arXiv:2306.02707.
- Chansung Park, Juyong Jiang, Fan Wang, Sayak Paul, and Jing Tang. 2024. [Llamaduo: Llmops pipeline for seamless migration from service llms to small-scale local llms](#). Preprint, arXiv:2408.13467.
- Ajay Patel, Colin Raffel, and Chris Callison-Burch. 2024. Datadreamer: A tool for synthetic data generation and reproducible llm workflows. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3781–3799.
- Baolin Peng, Chunyuan Li, Pengcheng He, Michel Galley, and Jianfeng Gao. 2023. Instruction tuning with gpt-4. *arXiv preprint arXiv:2304.03277*.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.
- David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R Bowman. 2024. Gpqa: A graduate-level google-proof q&a benchmark. In *First Conference on Language Modeling*.
- Gene Ruesamen. 2023. [GitHub - gururise/AlpacaDataCleaned: Alpaca dataset from Stanford, cleaned and curated](#) — github.com. <https://github.com/gururise/AlpacaDataCleaned>.
- Li Shen, Yan Sun, Zhiyuan Yu, Liang Ding, Xinmei Tian, and Dacheng Tao. 2024. On efficient training of

- large-scale deep learning models. *ACM Computing Surveys*, 57(3):1–36.
- Shivchander Sudalairaj, Abhishek Bhandwalder, Aldo Pareja, Kai Xu, David D. Cox, and Akash Srivastava. 2024. [Lab: Large-scale alignment for chatbots](#). Preprint, arXiv:2403.01081.
- Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Stanford alpaca: An instruction-following llama model. [https://github.com/tatsu-lab/stanford\\_alpaca](https://github.com/tatsu-lab/stanford_alpaca).
- Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).
- Ruida Wang, Wangchunshu Zhou, and Mrinmaya Sachan. 2023a. Let’s synthesize step by step: Iterative dataset synthesis with large language models by extrapolating errors from small models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 11817–11831.
- Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2023b. Self-instruct: Aligning language models with self-generated instructions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13484–13508.
- Jason Wei and Kai Zou. 2019. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6382–6388.
- John Wieting and Kevin Gimpel. 2018. Paranzmt-50m: Pushing the limits of paraphrastic sentence embeddings with millions of machine translations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 451–462.
- Can Xu, Qingfeng Sun, Kai Zheng, Xiubo Geng, Pu Zhao, Jiazhan Feng, Chongyang Tao, Qingwei Lin, and Daxin Jiang. 2024. [WizardLM: Empowering large pre-trained language models to follow complex instructions](#). In *The Twelfth International Conference on Learning Representations*.
- Yang Xu, Yongqiang Yao, Yufan Huang, Mengnan Qi, Maoquan Wang, Bin Gu, and Neel Sundaresan. 2023. [Rethinking the instruction quality: Lift is what you need](#). Preprint, arXiv:2312.11508.
- Zhangchen Xu, Fengqing Jiang, Luyao Niu, Yuntian Deng, Radha Poovendran, Yejin Choi, and Bill Yuchen Lin. 2025. [Magpie: Alignment data synthesis from scratch by prompting aligned LLMs with nothing](#). In *The Thirteenth International Conference on Learning Representations*.
- Zhaojian Yu, Xin Zhang, Ning Shang, Yangyu Huang, Can Xu, Yishujie Zhao, Wenxiang Hu, and Qiu Feng Yin. 2024. Wavecoder: Widespread and versatile enhancement for code large language models by instruction tuning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5140–5153.
- Rowan Zellers, Ari Holtzman, Yonatan Bisk, Ali Farhadi, and Yejin Choi. 2019. Hellaswag: Can a machine really finish your sentence? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4791–4800.
- Weihao Zeng, Can Xu, Yingxiu Zhao, Jian-Guang Lou, and Weizhu Chen. 2024. Automatic instruction evolving for large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6998–7018.
- Hao Zhao, Maksym Andriushchenko, Francesco Croce, and Nicolas Flammarion. 2024a. Long is more for alignment: a simple but tough-to-beat baseline for instruction fine-tuning. In *Proceedings of the 41st International Conference on Machine Learning*, pages 60674–60703.
- Yingxiu Zhao, Bowen Yu, Binyuan Hui, Haiyang Yu, Minghao Li, Fei Huang, Nevin L Zhang, and Yongbin Li. 2024b. Tree-instruct: A preliminary study of the intrinsic relationship between complexity and alignment. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16776–16789.
- Yaowei Zheng, Richong Zhang, Junhao Zhang, YeYanhao YeYanhao, and Zheyang Luo. 2024. Llamafactory: Unified efficient fine-tuning of 100+ language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 400–410.
- Chunting Zhou, Pengfei Liu, Puxin Xu, Srinivasan Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, Lili Yu, et al. 2023a. Lima: Less is more for alignment. *Advances in Neural Information Processing Systems*, 36:55006–55021.
- Daquan Zhou, Kai Wang, Jianyang Gu, Xiangyu Peng, Dongze Lian, Yifan Zhang, Yang You, and Jiashi Feng. 2023b. Dataset quantization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17205–17216.
- Haotian Zhou, Tingkai Liu, Qianli Ma, Yufeng Zhang, Jianbo Yuan, Pengfei Liu, Yang You, and Hongxia Yang. 2025. Davir: Data selection via implicit reward for large language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9220–9237.
- Jeffrey Zhou, Tianjian Lu, Swaroop Mishra, Siddhartha Brahma, Sujoy Basu, Yi Luan, Denny Zhou, and Le Hou. 2023c. [Instruction-following evaluation for large language models](#). Preprint, arXiv:2311.07911.

## A Computational Cost Analysis

We analyzed the computational cost of Mido’s optimization stages on 7B parameter models (LLaMA-3.1-8B, Mistral-7B-v0.3) using 50k-100k sample datasets (Alpaca, Alpaca-4o-mini, WizardLM) on  $8 \times$  NVIDIA A100 GPUs.

Each optimization iteration, encompassing data selection via *loss patterns*, *embedding cluster dynamics*, and *self-alignment scores*, followed by refinement, typically completes in under 30 minutes. This efficiency is largely due to the parallelizable nature of the diagnostic modules and the use of acceleration techniques: CUDA for neighbor computation in *Embedding Cluster Dynamics* and vLLM (Kwon et al., 2023) for batched inference during *Self-alignment Score* calculation. Table 4 provides a detailed time breakdown per component, underscoring Mido’s practical efficiency.

## B Hyperparameters Analysis

### B.1 The Impact of Neighbor Number

We also explore how the number of neighbors  $k$  used in the *Embedding Cluster Dynamics* affects the overall performance of Mido. By varying the number of neighbors, we analyze its impact on dataset diversity and model performance. Table 5 presents the results of this analysis. We find that the optimal number of neighbors is  $k = 2$ , which achieves the best balance between diversity and performance. This setting ensures that the dataset is sufficiently expanded to enhance model generalization while avoiding excessive noise that may degrade performance.

### B.2 The Impact of Iterations

As shown in Figure 7, we tested the number of iterations on the Alpaca dataset and found that the model’s performance significantly declined after the third iteration. Therefore, we chose to optimize each dataset for three iterations. This optimal number is not necessarily fixed and may vary depending on the threshold of each iteration.

### B.3 The Impact of Thresholds

The amount of data selected for refinement by each module (Loss Patterns (*Complexity*), Embedding Cluster Dynamics (*Diversity*), and Self-alignment Scores (*Quality*)) is governed by dynamic thresholds  $\tau = \mu + m\sigma$ , where  $\mu$  and  $\sigma$  are the mean and standard deviation of the respective signal values (loss, diversity score, quality score) across the

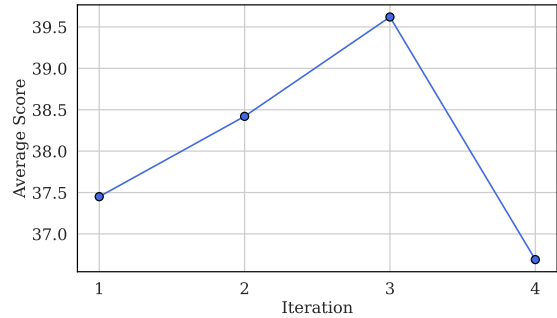


Figure 7: Performance trends on the Alpaca dataset across different iterations. The model’s performance peaks at three iterations and declines thereafter.

dataset. The multiplier  $m$  is a key hyperparameter that controls the stringency of these thresholds.

Our approach to setting  $m$  is guided by empirical analysis aimed at optimizing refinement effectiveness. Initial experiments (detailed in Section 5.3, Figure 3) indicated that refining a total unique proportion of approximately 10 – 20% of the dataset in the first iteration yields substantial performance improvements.

To determine appropriate  $m$  values for each module, we conducted a sensitivity analysis, presented in Table 7. This table shows how different combinations of  $m$  for complexity, diversity, and quality impact the total percentage of unique data selected for refinement and the resulting average model performance on benchmarks. The  $m$  values are varied (e.g., in 0.5 increments) for each signal, and combinations are chosen to target the 10-20% total selected data range. As shown, performance peaks when the selected data proportion falls within this empirically determined optimal range. For instance, the combination yielding 14.88% selected data achieved the best average score of 43.23. When multiple  $m$  combinations meet the 10 – 20% criterion, we select those with the smallest absolute  $m$  values (representing the mildest effective thresholds) that achieve this target, balancing refinement impact with efficiency.

The actual data sizes selected in each iteration for the experiments reported in the main paper, using  $m$  values derived from this sensitivity analysis (e.g., targeting the 15% mark initially), are detailed in Table 6. As the model’s performance improves over subsequent iterations, the amount of data flagged by these fixed  $m$  thresholds naturally decreases due to shifts in the signal distributions ( $\mu$  and  $\sigma$ ). This adaptive selection aligns with our observation that early training phases benefit from

| Method Component           | Time (Single A100 GPU)                   | Time ( $8 \times$ A100 GPUs, Data Parallelism) |
|----------------------------|------------------------------------------|------------------------------------------------|
| Loss patterns              | < 50 minutes                             | < 10 minutes                                   |
| Embedding cluster dynamics | < 40 minutes + neighbor computation time | < 10 minutes (CUDA acceleration)               |
| Self-alignment scores      | < 1 hour per metric (6 metrics)          | < 10 minutes (vLLM acceleration)               |

Table 4: Approximate overhead time cost of Middo’s optimization components per iteration. Timings are reported for processing datasets in the range of 50k-100k samples.

| $k$ | IFEval | GSM8K | MATH  | HumanEval | MBPP  | Hellaswag | ARC-c | Average      |
|-----|--------|-------|-------|-----------|-------|-----------|-------|--------------|
| 1   | 43.59  | 38.74 | 9.20  | 35.98     | 39.8  | 48.59     | 17.17 | 33.3         |
| 2   | 51.56  | 43.21 | 10.72 | 40.85     | 41.00 | 57.47     | 12.12 | <b>35.72</b> |
| 3   | 40.82  | 40.49 | 9.50  | 32.32     | 39.20 | 59.72     | 8.59  | 32.95        |

Table 5: Impact of the number of neighbors ( $k$ ) in the Embedding Cluster Dynamics on Middo performance. The table shows the performance across various benchmarks for different values of  $k$ , indicating that  $k = 2$  yields the best overall average score.

| Dataset                | iteration | loss      | neighbor   | self       | total  |
|------------------------|-----------|-----------|------------|------------|--------|
| <i>LLaMA-3.1-8B</i>    |           |           |            |            |        |
| Alpaca                 | init      | $m = 1$   | $m = -1$   | $m = -1.5$ | 52,002 |
|                        | iter1     | 1,180     | 1,924      | 1,159      | 53,939 |
|                        | iter2     | 299       | 1,853      | 108        | 55,811 |
|                        | iter3     | 242       | 1,822      | 381        | 57,636 |
| Alpaca 4o-mini         | init      | $m = 0$   | $m = -1$   | $m = -0.5$ | 52,002 |
|                        | iter1     | 5,684     | 8,032      | 4,145      | 60,865 |
|                        | iter2     | 611       | 2,291      | 876        | 63,184 |
|                        | iter3     | 472       | 2,127      | 661        | 65,324 |
| Wizard                 | init      | $m = 1$   | $m = -1.5$ | $m = -2$   | 70,000 |
|                        | iter1     | 3,585     | 3,585      | 2,690      | 73,642 |
|                        | iter2     | 959       | 3,341      | 1,016      | 76,993 |
|                        | iter3     | 751       | 3,414      | 420        | 80,419 |
| <i>Mistral-7B-v0.3</i> |           |           |            |            |        |
| Alpaca                 | init      | $m = 0.5$ | $m = -2$   | $m = -1$   | 52,002 |
|                        | iter1     | 2,418     | 2,111      | 2,367      | 54,131 |
|                        | iter2     | 1,985     | 2,091      | 932        | 56,268 |
|                        | iter3     | 1,788     | 1,982      | 352        | 58,348 |
| Alpaca 4o-mini         | init      | $m = 1$   | $m = -2$   | $m = -2.5$ | 52,002 |
|                        | iter1     | 1,407     | 7,691      | 1,499      | 59,696 |
|                        | iter2     | 1,278     | 9,116      | 1,045      | 68,874 |
|                        | iter3     | 1,346     | 2,487      | 661        | 74,036 |
| Wizard                 | init      | $m = 1$   | $m = -1.5$ | $m = -1.5$ | 70,000 |
|                        | iter1     | 5,637     | 5,709      | 5,258      | 76,429 |
|                        | iter2     | 3,558     | 5,999      | 6,310      | 82,501 |
|                        | iter3     | 3,885     | 6,229      | 3,767      | 89,178 |

Table 6: Data Size Details Across Iterative Refinement. For each dataset, the table lists the number of samples selected by the three components—*loss* (*Loss Patterns*), *neighbor* (*Embedding Cluster Dynamics*), and *self* (*Self-alignment Scores*). During each iteration, along with the total data size after refinement. The *init* row represents the original dataset size and the threshold controlling hyperparameter  $m$  corresponding to each component.

addressing a broader set of initial complexities and diversities, while later stages refine more nuanced aspects.

We do not place excessive emphasis on the improvements brought about by differences in data

| Complexity | Diversity  | Quality    | Total Selected | Percentage | Performance  |
|------------|------------|------------|----------------|------------|--------------|
| $m = 0$    | $m = -1$   | $m = -1.5$ | 15.8k          | 30.45%     | 41.81        |
| $m = 0.5$  | $m = -1.5$ | $m = -1.5$ | 7.7k           | 14.88%     | <b>43.23</b> |
| $m = 1$    | $m = -2$   | $m = -1.5$ | 4.3k           | 8.20%      | 41.96        |
| $m = 1.5$  | $m = -2.5$ | $m = -1.5$ | 2.6k           | 5.09%      | 41.55        |
| $m = 2$    | $m = -3$   | $m = -4$   | 1.3k           | 2.44%      | 40.87        |
| $m = 3$    | $m = -3.5$ | $m = -10$  | 0.5k           | 0.92%      | 39.64        |
| $m = 4$    | $m = -4$   | $m = -12$  | 0.1k           | 0.26%      | 38.69        |

Table 7: Sensitivity analysis for the threshold multiplier  $m$  on the Alpaca dataset (first iteration). The table shows the impact of varying  $m$  for complexity, diversity, and quality modules on the total unique data selected (sum and percentage) and the average model performance (mean score across benchmarks).

volume, so our selection may not necessarily be optimal.

## C Experimental Details

### C.1 Instruction Fine-tune Dataset

We evaluate Middo on three general instruction fine-tuning datasets.

- **Alpaca** (Taori et al., 2023): consists of 52,002 instruction-response pairs generated by Stanford University using the self-instruct (Wang et al., 2023b) method based on OpenAI’s text-davinci-003. This dataset is designed for fine-tuning dialogue models similar to ChatGPT to achieve efficient instruction-following capabilities.
- **Alpaca-4o-mini**: to evaluate performance on a higher-quality response dataset, we generated responses for all Alpaca instructions using GPT-4o mini, creating the Alpaca-4o-mini dataset.
- **WizardLM** (Xu et al., 2024): 70K data generated based on Evol-Instruct, which aims

to generate more complex instruction data through a recursive evolutionary approach in order to improve the model's reasoning and instruction comprehension.

## C.2 Models

We primarily conducted experiments on LLaMA 3.1-8B, and additionally performed extra experiments on Mistral 7B-v0.3.

- **LLaMA 3.1-8B** (Dubey et al., 2024): LLaMA 3.1-8B is a large language model released by Meta, featuring 8 billion (8B) parameters. It is part of the LLaMA (Large Language Model Meta AI) series, focusing on efficient reasoning and text generation capabilities. LLaMA 3.1-8B excels in code generation, language understanding, and conversational tasks, optimizing inference speed and training efficiency, making it suitable for research, commercial applications, and AI studies.
- **Mistral 7B-v0.3** (Jiang et al., 2023): Mistral 7B-v0.3 is an open-source language model developed by Mistral AI, featuring 7 billion parameters. It is optimized based on the Transformer architecture, emphasizing efficiency and multitasking capabilities. Compared to earlier versions, this model shows improvements in coding, mathematics, and reasoning tasks, making it suitable for chatbots, programming assistance, and natural language processing tasks. Mistral 7B-v0.3 incorporates feedback from the open-source community to enhance inference efficiency, delivering high performance with reduced computational resources.
- **MMLU (Massive Multitask Language Understanding)** (Hendrycks et al., 2021b): a large-scale, multi-task language understanding benchmark that covers 57 subjects, testing models on their knowledge and reasoning abilities across fields such as history, law, mathematics, and medicine. It serves as a significant indicator of general artificial intelligence knowledge levels.
- **GSM8K (Grade School Math 8K)** (Cobbe et al., 2021): a dataset specifically created for solving mathematical problems, containing approximately 8,500 elementary school math questions that primarily focus on basic arithmetic, logical reasoning, and text comprehension skills. This dataset is used to evaluate models' mathematical computation and reasoning abilities.
- **MATH** (Hendrycks et al., 2021a): consists of math competition problems from middle school and college levels, covering areas such as algebra, geometry, number theory, and calculus. This dataset is more challenging than GSM8K and is primarily used to assess models' performance on advanced mathematical reasoning tasks.
- **HumanEval** (Chen et al., 2021): a dataset for evaluating code generation capabilities, featuring a series of Python programming problems, each with a clear function signature and test cases. This dataset is commonly used to measure AI performance in automated code generation and programming tasks.
- **MBPP (Mostly Basic Programming Problems)** (Austin et al., 2021): a benchmark dataset for code generation, containing 1,000 basic programming questions that cover data structures, algorithms, and logical reasoning. It is suitable for assessing AI capabilities in fundamental programming tasks.
- **Hellaswag** (Zellers et al., 2019): a benchmark dataset for commonsense reasoning, consisting of a series of incomplete sentences that require models to select the most reasonable ending. This dataset tests models' contextual understanding and reasoning abilities by designing misleading options.

## C.3 Benchmarks

We assess model performance on general knowledge, mathematical problem-solving, code generation and commonsense reasoning benchmarks.

- **IFEval (Instruction Following Evaluation)** (Zhou et al., 2023c): a benchmark dataset designed to assess the instruction-following capabilities of large models. It encompasses various tasks, including general knowledge question answering, commonsense reasoning, and mathematical reasoning, aiming to measure the understanding and accuracy of language models when executing complex instructions.

- **GPQA (Graduate-Level Google-Proof Q&A)** (Rein et al., 2024): a challenging dataset designed to evaluate the capabilities of LLMs and scalable oversight mechanisms. Let me provide more details about it.

#### C.4 Baselines

We compare Middo with both existing data selection and data augmentation methods on the Alpaca dataset.

##### Data Selection Methods.

- **Alpaca-clean** (Ruebsamen, 2023): a cleaned version of the Alpaca dataset that removes low-quality samples and duplicates, aiming to improve the overall quality of the dataset.
- **Superfiltering** (Li et al., 2024b): using smaller, weaker language models (such as GPT-2) as data filters to compute IFD allows for the selection of high-quality instruction tuning data.
- **Long** (Zhao et al., 2024a): directly select the 1,000 samples with the longest responses as training data.
- **AlpaGasus** (Chen et al., 2024): utilize powerful LLMs (such as ChatGPT) to automatically assess the sample quality in the Alpaca dataset and filter out high-quality data to enhance model training effectiveness.

##### Data Augmentation Methods.

- **Alpaca-GPT4** (Peng et al., 2023): a data augmentation method that uses GPT-4 to generate additional training data for the Alpaca dataset.
- **I-SHEEP** (Liang et al., 2025): a data augmentation method that uses a self-supervised learning approach to generate additional training data for the Alpaca dataset.
- **WizardLM** (Xu et al., 2024): 70K data generated based on Evol-Instruct, which aims to generate more complex instruction data through a recursive evolutionary approach in order to improve the model’s reasoning and instruction comprehension.

| Hyperparameter              | Value              |
|-----------------------------|--------------------|
| <i>LLaMA-3.1-8B</i>         |                    |
| Learning Rate               | $2 \times 10^{-5}$ |
| Number of Epochs            | 1                  |
| Number of Devices           | 8                  |
| Per-device Batch Size       | 4                  |
| Gradient Accumulation Steps | 8                  |
| Learning Rate Scheduler     | cosine             |
| Warmup Ratio                | 0.03               |
| Max Sequence Length         | 4096               |
| <i>Mistral-7B-v0.3</i>      |                    |
| Learning Rate               | $1 \times 10^{-5}$ |
| Number of Epochs            | 1                  |
| Number of Devices           | 8                  |
| Per-device Batch Size       | 4                  |
| Gradient Accumulation Steps | 8                  |
| Learning Rate Scheduler     | cosine             |
| Warmup Ratio                | 0.03               |
| Max Sequence Length         | 4096               |

Table 8: Hyperparameters used for fine-tuning.

#### C.5 Hyperparameters

**Fine-tune.** For LLaMA-3.1-8B, we follow the Alpaca GitHub repository<sup>1</sup>, setting the batch size to 32, the learning rate to  $2 \times 10^{-5}$ , and the warmup ratio to 0.03. For Mistral-7B-v0.3, we adjust the learning rate to  $1 \times 10^{-5}$ , as per official recommendations<sup>2</sup>. All the hyperparameters are detailed in Table 8.

**Data Synthetic.** We use the OpenAI API to generate data by GPT-4o-mini, setting both temperature and top\_p to 1.0 to guarantee diversity.

**Evaluation.** All benchmarks are conducted in zero-shot and we conducted the tests using the default configuration of OpenCompass. All the hyperparameters are detailed in Table 9.

All experiments are conducted on  $8 \times$  NVIDIA Tesla A100 GPUs about 50 GPU hours.

#### D Self-alignment Scores

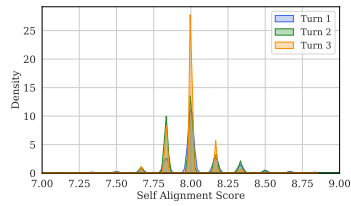
We provide detailed self-alignment score evolution across iterations on the Alpaca, Alpaca-4o-mini, and WizardLM datasets in Figure 8. These figures illustrate the dynamic evolution of self-alignment

<sup>1</sup>[https://github.com/tatsu-lab/stanford\\_alpaca](https://github.com/tatsu-lab/stanford_alpaca)

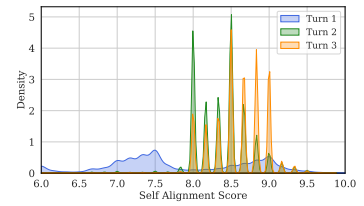
<sup>2</sup><https://docs.mistral.ai/capabilities/finetuning>



(a) Alpaca dataset.



(b) Alpaca-4o-mini dataset.



(c) WizardLM dataset.

Figure 8: Self-alignment score evolution across iterations. The x-axis represents the self-alignment scores, while the y-axis shows the density of data points.

| Hyperparameter     | Value   |
|--------------------|---------|
| Pass@n             | $n = 1$ |
| Presence Penalty   | 0.0     |
| Frequency Penalty  | 0.0     |
| Repetition Penalty | 1.0     |
| Temperature        | 0.0     |
| Top_p              | 1.0     |
| Top_k              | -1      |
| Min_p              | 0.0     |
| Max Tokens         | 4096    |
| Min Tokens         | 0       |

Table 9: Hyperparameters used for evaluation.

scores across iterations, highlighting the continuous improvement in dataset quality and alignment with model capabilities.

## E Case Study

| Self-Alignment Instruction Score Prompt Template                                                                                                                                                                                                                                                                                                                                                                                                        |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>We would like to request your feedback on the {dimension} of the prompt displayed below.</p> <p>Prompt:<br/>Give three tips for staying healthy.</p> <p>Please rate according to the {dimension} of the prompt to evaluate {explanations}. Each prompt is scored on a scale of 0 to 10, with higher scores indicating higher {dimension}. Try to avoid scoring a full 10. Give your rating number first, then give a explanation of your rating.</p> |
| Self-Alignment Instruction Score Dimensions and Explanations                                                                                                                                                                                                                                                                                                                                                                                            |
| <p>"factuality": "whether the information provided in the prompt is accurate and based on reliable facts and data",</p> <p>"clarity": "whether the prompt is clear and understandable, and whether it uses concise language and structure",</p> <p>"completeness": "whether the prompt provides sufficient information and details".</p>                                                                                                                |
| Instruction Clarity Score Example                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <p>7. The prompt is clear and understandable, but it could be more concise.</p>                                                                                                                                                                                                                                                                                                                                                                         |

Figure 9: Self-Alignment instruction score example.

| Self-Alignment Response Score Prompt Template                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>We would like to request your feedback on the {dimension} of the prompt displayed below.</p> <p>Prompt:<br/>What are the three primary colors?</p> <p>Response:<br/>The three primary colors are red, blue, and yellow.</p> <p>Please rate according to the {dimension} of the response to evaluate {explain}. Each response is scored on a scale of 0 to 10, with higher scores indicating higher {dimension}. Try to avoid scoring a full 10. Give your rating number first, then give a explanation of your rating.</p> |
| Self-Alignment Response Score Dimensions and Explanations                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <p>"factuality": "whether the information provided in the response is accurate and based on reliable facts and data",</p> <p>"clarity": "whether the response is clear and understandable, and whether it uses concise language and structure",</p> <p>"completeness": "whether the response provides sufficient information and details".</p>                                                                                                                                                                                |
| Response Clarity Score Example                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <p>8.5. The response is clear and understandable, but it could be more concise.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                           |

Figure 10: Self-Alignment response score example.

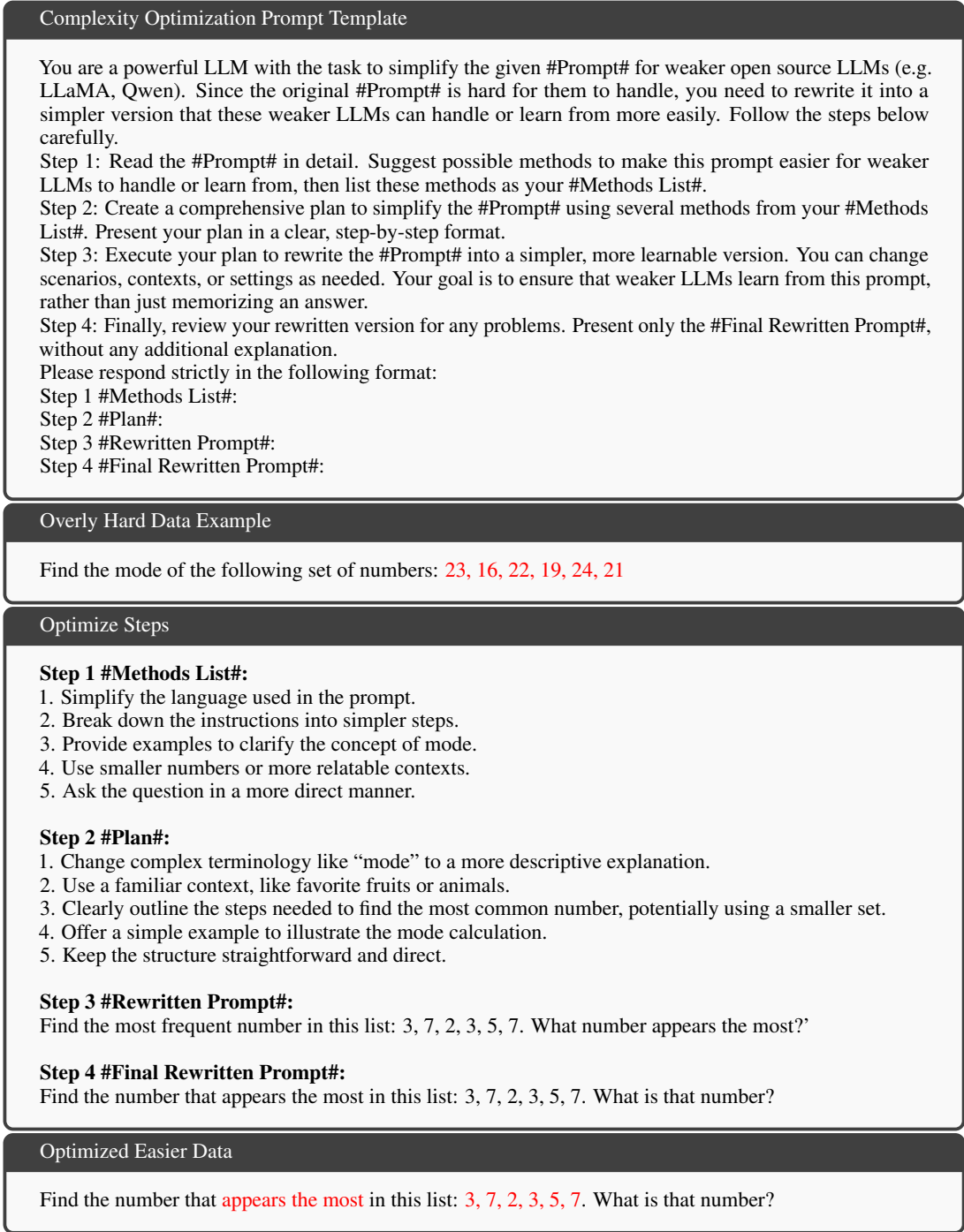


Figure 11: Complexity optimization example.

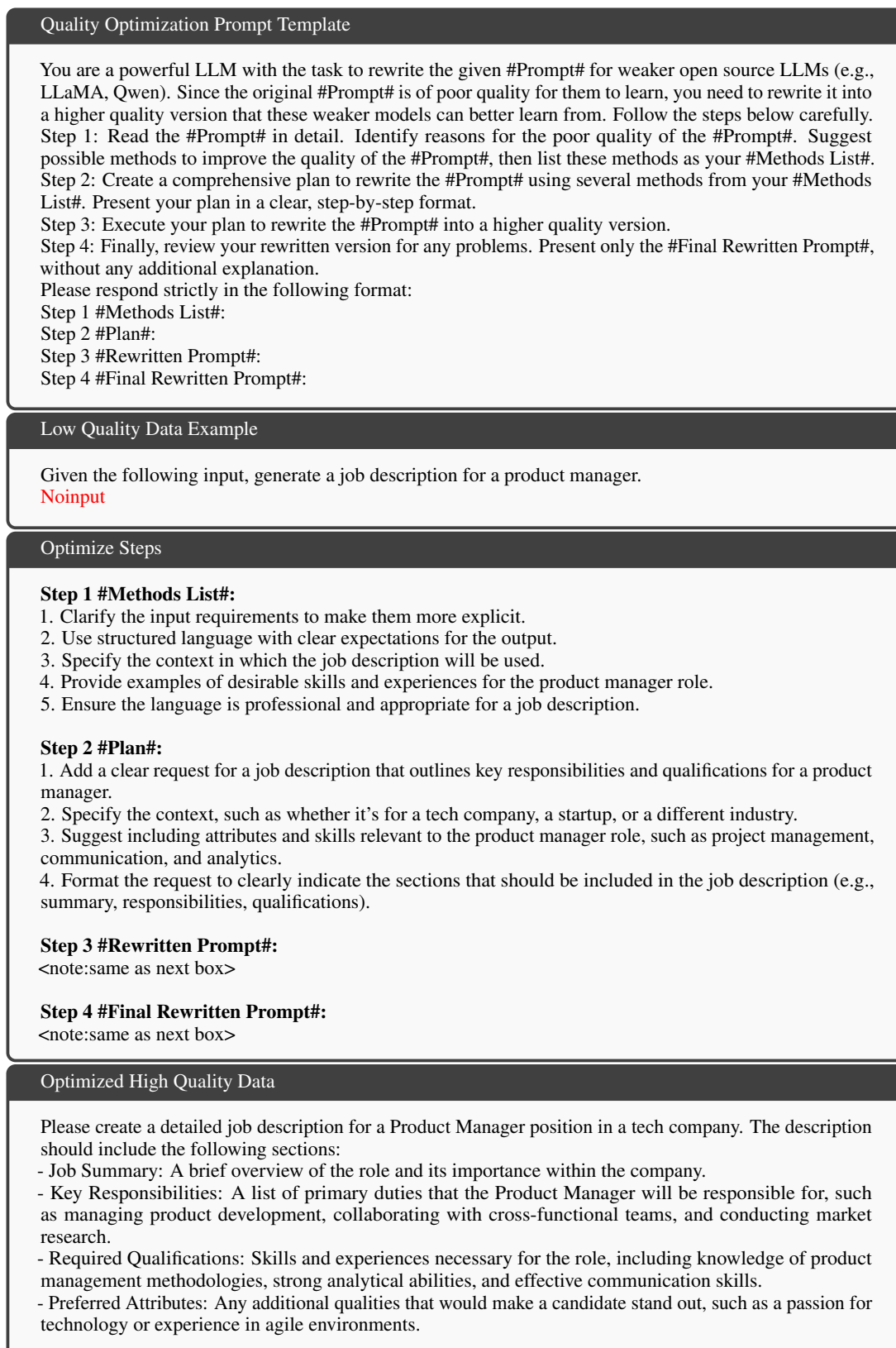


Figure 12: Quality optimization example.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Diversity Extension Prompt Template                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <p>You are a powerful LLM with the task to create brand new prompts for weaker open source LLMs (e.g. LLaMA, Qwen). You need to create a brand new complete prompt for them to learn in order to improve their knowledge and skills. Follow the steps below carefully.</p> <p>Use #Hint Prompt 1# and #Hint Prompt 2# as guiding examples. Then read the #Core Prompt# in detail. Be inspired to suggest additional new prompts, and ultimately create only one completely original and diverse #New Prompt#.</p> <p>Please respond strictly in the following format:</p> <p>#New Prompt#:"</p> |
| Sparse Data And Neighbors                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <p><b>#Hint Prompt 1#:</b><br/>How long did Shakespeare live?</p> <p><b>#Hint Prompt 2#:</b><br/>How did the Industrial Revolution change society?</p> <p><b>#Core Prompt#:</b><br/>How did Julius Caesar die?</p>                                                                                                                                                                                                                                                                                                                                                                              |
| Extensioned Data                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <p>What were the key factors that led to the fall of the Roman Empire?</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |

Figure 13: Diversity extension example.