# WisdoMentor: An Advanced Tutoring Large Language Model for Enhanced AI Educational Support

**Jinghao Lin**
linjinghao01@baidu.com

**Hao Wang**
email@domain

Guoqing Chen
email@domain

Yi Yan
email@domain

## Abstract

In the rapidly evolving field of artificial intelligence (AI), students face the challenge of navigating complex concepts and acquiring comprehensive knowledge. To address this, we present WisdoMentor, an advanced AI model specifically designed to provide academic support in the field of AI.

Motivated by the increasing demand for effective educational resources, our model incorporates innovative techniques like dataset partitioning, student-friendly instruction generation, feedback preference alignment, and RAG-based answer confidence enhancement. These techniques aim to enhance the learning experience and bridge the gap between theoretical concepts and practical applications.

To evaluate the proficiency of WisdoMentor, we introduce the AI Mentor Benchmark, a novel evaluation framework that assesses the model's capabilities comprehensively. Through rigorous performance tests, we demonstrate the superiority of WisdoMentor over other general-purpose models in terms of accuracy and effectiveness.

We sincerely invite researchers to explore the code and resources of WisdoMentor on GitHub at https://github.com/linjh1118/WisdoMentor.

## 1 Introduction

In the realm of AI-assisted learning, the advent of Large Language Models (LLMs) has catalyzed a paradigm shift, promising to redefine the landscape of education through innovative applications of artificial intelligence. Among these, WisdoMentor stands out as a beacon of advancement, offering a structured framework for personalized learning experiences that bridge the gap between learners and educational content. At its core, WisdoMentor is built upon a meticulously crafted architecture that encompasses three key components: pre-training, supervised fine-tuning, and reinforcement learning from human feedback. This comprehensive structure not only enables the model to ingest and comprehend vast amounts of educational data but also facilitates continuous refinement based on real-time interactions with users. The pre-training phase lays the foundation for WisdoMentor, exposing it to a diverse corpus of educational materials spanning various subjects and domains. Through this process, the model acquires a broad understanding of language and knowledge, providing a robust base upon which subsequent learning and adaptation can occur. Supervised fine-tuning represents the next stage in WisdoMentor's evolution, wherein the model is tailored to specific educational contexts through targeted training on domain-specific datasets. This process allows the model to refine its understanding of key concepts, terminology, and learning objectives, thereby enhancing its ability to generate relevant and accurate responses in educational settings. However, WisdoMentor goes beyond mere adaptation; it actively engages in a continuous cycle of learning and improvement through reinforcement learning from human feedback. By soliciting input from users and incorporating human preferences and evaluations into its training regimen, the model iteratively refines its responses, ensuring that they align closely with user expectations and learning goals. In this paper, we present a detailed exploration of WisdoMentor's structure and functionality, highlighting the significance of each component in facilitating AI-assisted learning. We describe the model's pre-training methodology, supervised fine-tuning process, and reinforcement learning framework, shedding light on the intricacies of its design and implementation. Furthermore, we present the results of experiments conducted to evaluate the performance of WisdoMentor across various educational tasks and scenarios. Through rigorous testing and analysis, we demonstrate the model's efficacy in supporting personalized learning experiences, generating informative and engaging content, and fostering meaningful interactions

between learners and educational resources.

## 2 Related Works

In the field of natural language processing (NLP), general large language models (LLMs) have made significant strides, with models like ChatGPT (OpenAI, 2022) and Qwen (Bai et al., 2023) exemplifying the state-of-the-art. ChatGPT, crafted by OpenAI, has become a paragon of dialogue-based AI, utilizing its transformer architecture to process extensive data and produce contextually relevant responses, thereby setting a new standard for conversational AI. Concurrently, Qwen, a model tailored for the Chinese language, has demonstrated adeptness in handling the intricacies of Chinese NLP tasks, including text generation and comprehension, marking a significant advancement in the domain. These models not only epitomize the current achievements in NLP but also pave the way for future innovations that will further enhance AI's ability to understand and generate human-like language.

The emergence of specialized LLMs such as Xuanyuan (Zhang and Yang, 2023), SongComposer (Ding et al., 2024), and ChatLaw (Cui et al., 2023) underscores a common theme: the application of large-scale models in specific domains beyond general-purpose NLP. These models leverage the power of deep learning to provide sophisticated interactions and insights within their respective areas, be it legal consultation, music composition, or other specialized knowledge bases.

Continual Pre-training (CPT) is an essential approach for keeping LLMs up-to-date with the ever-evolving human knowledge and linguistic patterns. Initial CPT methodologies prioritized domain adaptation to retain knowledge from prior tasks (Gururangan et al., 2020). Subsequent advancements introduced techniques like logit distillation, which effectively maintained foundational knowledge while incorporating new data (Jin et al., 2021). The seminal work by Gupta et al. (Gupta et al., 2023)on re-warming strategies optimized learning rates, thereby improving downstream task performance without forsaking previously acquired knowledge. Most recently, the Llama-Pro model employs a block expansion technique that efficiently assimilates new domain-specific knowledge and preserves the model's initial capabilities through targeted post-training (Wu et al., 2024). These developments collectively refine the CPT process, enabling LLMs to adapt to novel data with improved efficiency and effectiveness.

In the landscape of recent large model developments, data optimization has emerged as a cornerstone for enhancing model performance. For instance, the Dolma dataset, which was created by integrating data from seven different sources, has demonstrated the effectiveness of combining diverse data for training large-scale models (Soldaini et al., 2024). Additionally, the use of data augmentation techniques, such as those implemented in the GPT-4 model, has been shown to significantly improve the robustness and generalizability of the model by expanding the diversity of the training data (Brown et al., 2020). Furthermore, the application of advanced data filtering methods, like those based on the Hugging Face's Transformers library[1], allows for the removal of noisy or irrelevant data, thereby refining the dataset and leading to more accurate model predictions. These approaches underscore the importance of a well-optimized dataset in the development of high-performing large models.

Supervised Fine-Tuning (SFT) is a process where a pre-trained model is further trained on a specific task with labeled data to improve its performance and accuracy for the target task. Traditional SFT methods involve continuing the training process on a task-specific dataset, optimizing the model's parameters for the target task's performance (Devlin et al., 2018). More recent developments have explored innovative approaches such as Self-Instruct tuning, which leverages an iterative bootstrapping process to generate instruction-following data and significantly improves the model's ability to adhere to human-provided instructions (Wang et al., 2022). This method has sparked interest in the research community due to its potential to create diverse and creative instruction-following datasets with minimal human annotation. Additionally, research has focused on enhancing SFT through techniques like Contrastive Initialization (COIN), which aims to improve the semantic relation among instances by enriching the semantic information captured by self-supervised models (Pan et al., 2023).

Direct Preference Optimization (DPO) has been a pivotal approach in fine-tuning large language models to align with human preferences, offering a simpler and more stable alternative to complex reinforcement learning methods. While traditional

[1] https://huggingface.co/transformers/

DPO effectively uses preference data to optimize model outputs, it often operates in a single phase, which may not fully exploit the potential of incremental learning (Rafailov et al., 2023). Stepwise DPO (sDPO) introduces a novel extension by dividing the preference datasets and utilizing them incrementally (Kim et al., 2024). This phased approach allows for the use of more precisely aligned reference models within the training framework,and has been demonstrated to outperform other models with more parameters.

## 3 Pretraining

### 3.1 Data

The scale of data has emerged as a pivotal determinant in the development of robust large language models. Crafting an efficacious pretraining dataset necessitates meticulous attention to ensuring diversity, encompassing a broad spectrum of genres, domains, and tasks. Our dataset is meticulously curated to fulfill these prerequisites, comprising publicly available web documents, encyclopedias, books, and source code, among others. Furthermore, it boasts multilingual coverage, with a substantial corpus in both English and Chinese.

To guarantee the quality of our pre-training data, we devised a meticulous data pre-processing protocol. For publicly available web data, we extract text from HTML sources and employ language identification tools for language determination. To augment data diversity, we leverage deduplication technologies, incorporating both exact match deduplication post-normalization and fuzzy deduplication utilizing MinHash and LSH algorithms. In order to weed out low-quality data, we amalgamate rule-based and machine learning-based methodologies. Specifically, we deploy a diverse array of models for content scoring, encompassing language models, text quality scoring models, and models for identifying potentially offensive or inappropriate content. Furthermore, we conduct manual sampling of text from diverse sources, subjecting it to rigorous review to ensure its integrity. To further enhance data quality, we selectively upsample data from specific sources, ensuring our models are trained on a varied corpus of high-caliber content.

### 3.2 Training Strategy

Inspired by *phi-2*, we employ a comprehensive range of metrics to effectively divide the dataset into distinct subsets. This segmentation strategy allows us to seamlessly integrate our innovative *Wisdo-WSD* scheduler into the training process. By incrementally introducing carefully selected data batches at different stages, our scheduler facilitates a gradual and precise acquisition of the dataset's underlying patterns by the model.

This dynamic approach to dataset inclusion not only promotes a more fine-tuned learning experience but also empowers the model to adapt and evolve as it gains deeper insights into the intricacies of the data. Through this iterative process, the model becomes increasingly adept at uncovering hidden relationships, correlations, and trends within the dataset, ultimately leading to enhanced accuracy and more robust predictions.

Furthermore, the *Wisdo-WSD* scheduler acts as a strategic guide, intelligently prioritizing and scheduling the introduction of specific data subsets based on their relevance and informative value. This ensures that the model receives a diverse range of examples, avoiding bias and allowing it to generalize better to unseen data.

By combining the power of data partitioning based on meaningful metrics with our *Wisdo-WSD* scheduler, we establish a powerful framework for progressive learning. This framework not only enables the model to acquire a deeper understanding of the dataset's intrinsic rules but also facilitates continuous improvement and adaptability, positioning our approach at the forefront of cutting-edge research in the field.

### 3.3 Tokenizer Extension

In this section, we present our approach to enhancing vocabulary and expertise in AI discussions. By analyzing highly cited papers and popular blogs, we have developed a methodology to expand the lexicon and improve decoding efficiency.

To expand the vocabulary, we have curated an AI-specific lexicon from key terms and concepts found in influential papers and blogs. We have also employed advanced NLP techniques, like word embeddings and contextual word representations, to capture contextual associations and enable accurate vocabulary choices.

To improve expertise, we have built an AI-driven knowledge enhancement framework. It utilizes cutting-edge techniques in knowledge graph construction and representation learning. By incorporating diverse data sources and advanced graph embedding algorithms, we encode semantic relationships in an AI-specific knowledge graph.

To facilitate knowledge retrieval, we have integrated a semantic search engine into our framework. It provides context-aware information from the knowledge graph, allowing researchers and enthusiasts to access up-to-date findings and expert insights.

These vocabulary enhancement and expertise improvement approaches will be integrated into the arXiv article, enabling a more comprehensive and insightful discussion on AI advancements. By embracing these advancements, we anticipate fostering more nuanced discussions, accelerating research progress, and enhancing collaboration in the field of AI.

## 4 Supervised Fine-tuning

Generation of Academic Instruction Following Capability using Gate Mechanisms In this chapter, we present our approach to enhancing the capability of generating academic instructions by leveraging various cutting-edge instruction generation techniques through gate mechanisms. Through the integration of these techniques, we have generated a significant number of synthetic academic instructions (referred to as SFT) and conducted comprehensive training to improve the ability to follow such instructions.

### 4.1 Gate Mechanisms for Instruction Generation

To generate high-quality academic instructions, we have incorporated gate mechanisms into our instruction generation framework. Gate mechanisms allow for the selective integration of different instruction generation techniques, enabling a more refined and diverse output.

We have explored state-of-the-art instruction generation techniques, including neural machine translation, sequence-to-sequence models, and transformer-based architectures. By utilizing gate mechanisms, we can dynamically adjust the contribution of each technique based on their performance and relevance to the academic context.

### 4.2 Generation of Synthetic Academic Instructions

To facilitate training and evaluation, we have generated a large corpus of synthetic academic instructions, referred to as SFT. These instructions cover a wide range of academic topics and are designed to mimic the complexity and diversity of real-world academic instructions.

By employing the gate mechanisms and integrating various instruction generation techniques, we have ensured that the SFT corpus captures the nuances and intricacies of academic language, including specific terminologies, structural patterns, and rhetorical devices commonly found in scholarly discourse.

## 5 Wisdo-MoE-DPO

In this chapter, we extend the concept of integrating multiple experts to the domain of preference alignment in DPO approach. While the traditional approach focuses on incorporating additional experts to enrich knowledge and improve performance, we propose a novel framework that introduces a general user layer to align user preferences. By doing so, we aim to enhance the personalization capabilities of DPO approach.

In DPO approach, the ultimate goal is to deliver personalized experiences that align with individual user preferences. However, understanding and capturing these preferences accurately can be challenging. Users often exhibit diverse and evolving preferences, making it crucial for DPO approach to adapt and align their recommendations accordingly.

As shown in Figure **??**, DPO skips the training step of the reward model based on the preference data pairs, and directly optimizes the final large model.

### 5.1 General User Layer

To address the challenge of preference alignment, we introduce a general user layer into the DPO framework. The general user layer represents the preferences and characteristics of the average user within the target user population. This layer serves as a reference point for aligning individual user preferences and acts as a baseline for comparison.

By incorporating the general user layer, we enable DPO approach to identify deviations or discrepancies between individual user preferences and the average preferences. This allows for targeted adjustments and personalized recommendations that cater to the unique preferences of each user.

### 5.2 DPO with General User Layer

Within the general user layer, we focus on adjusting and fine-tuning the user preferences to align them with the desired outcomes. This adjustment process involves leveraging various techniques, such

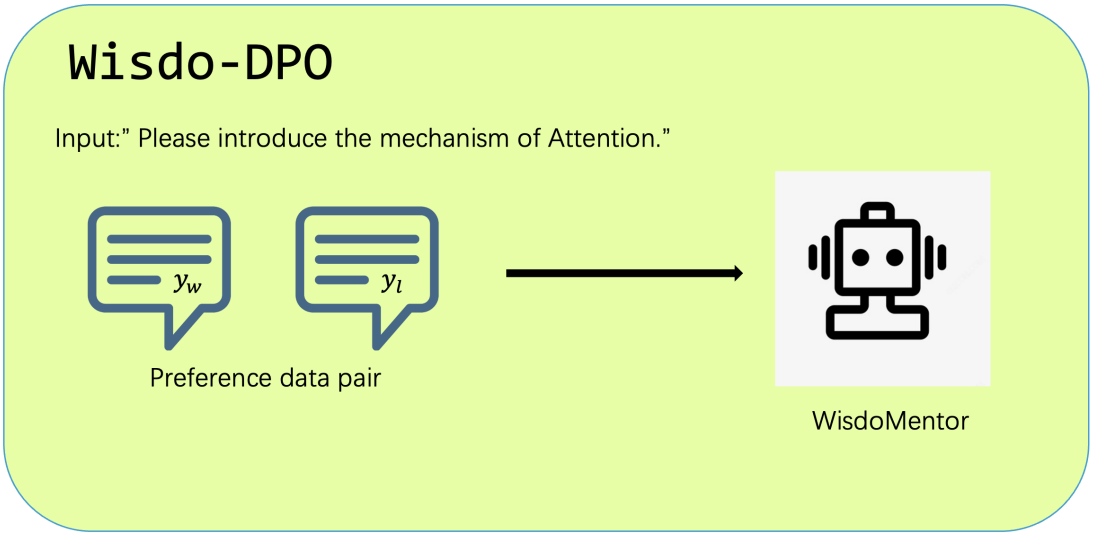


Figure 1: The model architecture is as shown in the figure. The FFN layer is divided into multiple expert layers, and a router decides to send the token to the most appropriate expert for processing.

as reinforcement learning, contextual bandits, or collaborative filtering, to iteratively refine the preferences.

By selectively adjusting the preferences of the general user layer, we ensure that the system captures the collective user preferences while adapting to individual variations. This approach strikes a balance between maintaining the overall coherence of recommendations and catering to the specific preferences of each user.

### 5.3 Mixture of Experts

Model size is one of the key factors to improve model performance. With a limited computing resource budget, training a larger model with fewer training steps often performs better than training a smaller model with more steps.

A significant advantage of Mixture of Experts(MoE) is their ability to be efficiently pretrained with far fewer computational resources than dense models require. This means you can significantly increase the size of your model or dataset within the same compute budget. Especially in the pre-training phase, hybrid expert models are often able to reach the same quality level faster than dense models.

Therefore, we introduced the MoE structure in WisdoMentor, that is, the router decides which token to send to which expert for processing, thereby obtaining better results.The model structure is shown in Figure **??**

## 6 Evalutions

In our experiment, our primary objective was to rigorously test and validate the outstanding performance of our model WisdoMentor by subjecting it to strict evaluation against both traditional benchmarks and benchmarks specifically tailored for assessing the proficiency of AI assistants.

### 6.1 Common Benchmark

We utilized the open-source tool UltraEval for comprehensive evaluation purposes. UltraEval serves as a foundational model capability evaluation framework, designed initially to offer a lightweight, user-friendly evaluation system. Its primary aim is to support the performance assessment of prominent large-scale models while meeting the rapid evaluation needs of model training teams. At its core, the framework leverages the vLLM open-source framework for reasoning and acceleration, thus ensuring both efficiency and precision throughout the evaluation process.

Regarding dataset selection, we meticulously curated a range of widely recognized authoritative datasets to ensure the thoroughness and reliability of our evaluation outcomes. These datasets span various domains and languages, encompassing, but not limited to:

- MMLU (Massive Multitask Language Understanding): Evaluate language model performance across 57 tasks, including 6 medical-

Table 1: Open source benchmark result, WisdoMentor's performance on nine major benchmark datasets, and comparison with current mainstream large models and a series of open source models.

| | MMLU | CMMLU | C-Eval | HumanEval | MBPP | GSM8K | MATH | ARC-E | ARC-C | BBH |
|---|---|---|---|---|---|---|---|---|---|---|
| Llama2-7B | 44.32 | 31.11 | 32.42 | 12.2 | 27.17 | 13.57 | 1.8 | 75.25 | 42.75 | 33.23 |
| Llama2-13B | 54.71 | 37.06 | 37.32 | 17.07 | 32.55 | 21.15 | 2.25 | 78.87 | 58.19 | 37.92 |
| Qwen-1.8B | 43.37 | 45.32 | 49.81 | 7.93 | 17.8 | 19.26 | 2.42 | 63.97 | 43.69 | 29.07 |
| Qwen-7B | 57.65 | 60.35 | 58.96 | 17.07 | 42.15 | 41.24 | 5.34 | 83.42 | 64.76 | 37.75 |
| Phi-2(2B) | 52.66 | 24.18 | 23.37 | 47.56 | **55.04** | **57.16** | 3.5 | **86.11** | 71.25 | **43.39** |
| ChatGLM2-6B | 45.77 | 49.21 | 52.05 | 10.37 | 9.38 | 22.74 | 5.96 | 74.45 | 56.82 | 32.6 |
| Baichuan2-7B-Chat | 53 | 53.5 | 53.28 | 21.34 | 32.32 | 25.25 | 6.32 | 79.63 | 60.15 | 37.46 |
| MiniCPM-2B | 53.46 | 51.07 | 51.13 | **50.00** | 47.31 | 53.83 | **10.24** | 85.44 | **68.00** | 36.87 |
| **WisdoMentor** | **60.64** | **69.19** | **71.18** | 2.44 | 45.43 | 50.34 | 7.81 | 84.60 | 67.74 | 40.14 |

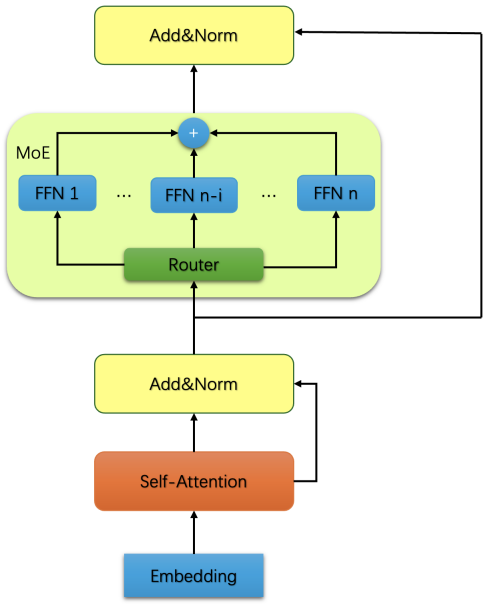


Figure 2: The model architecture is as shown in the figure. The FFN layer is divided into multiple expert layers, and a router decides to send the token to the most appropriate expert for processing.

related fields. Contains 15908 questions, with 123 verification questions and 1089 test questions in clinical topics.

- CMMLU (Chinese Multitask Language Understanding): Assess Chinese language and cultural knowledge with 67 topics and 11,528 questions, with at least 105 questions per subject.
- C-Eval: Evaluate model understanding and adaptability to Chinese knowledge with 13,948 multiple-choice questions across 4 major subject categories and 52 sub-categories.
- HumanEval: Assess code generation ability with 164 Python programming questions, including function headers, bodies, and unit tests, graded by Pass@k score.
- MBPP (Mostly Basic Programming Problems): Evaluate performance on 974 basic Python programming tasks to improve programming models.
- GSM8K (Grade School Math 8K): Solve 8,500 primary school mathematics vocabulary problems to support multi-step reasoning.
- MATH: Contain 12,500 math problems from competitions like AMC 10, AMC 12, AIME, with 7500 for training and 5000 for testing.
- ARC (Artificial Reading Comprehension): Contains about 8,000 scientific questions divided into ARC-Challenge and ARC-Easy subsets with a corpus of 14M scientific facts.
- BBH (BIG-Bench Hard): Suite of 23 challenging tasks in BIG-Bench not surpassed by previous language models.

By strategically selecting widely recognized datasets within each domain, we aimed to provide a comprehensive assessment of WisdoMentor's capabilities.

### 6.2 AI Mentor Benchmark

To thoroughly assess the performance and accuracy of WisdoMentor in the domain of paper question and answer, we meticulously crafted a comprehensive evaluation plan. This initiative centers on a meticulously tailored dataset derived from a selection of representative academic papers. The primary aim is to conduct an exhaustive evaluation of WisdoMentor's capacity to comprehend, analyze, and respond to queries pertaining to academic literature.

In the initial phase of dataset creation, we meticulously curated 20 papers from various scholarly repositories. These papers not only wield substantial influence within academia but also span diverse verticals. Through meticulous scrutiny of these papers, we distilled crucial and foundational information from each, encompassing research context, objectives, methodologies, results, as well as the authors' principal viewpoints and arguments.

Following this, we formulated four pivotal questions for each paper, aimed at delving deeply into the core concepts and resolutions posited within. We ensured that these questions not only scrutinize the model's grasp of the paper's minutiae but also assess its comprehension of the overarching structure and rationale. To ensure impartial and precise evaluation, we furnished expert-validated answers to each question, encompassing all requisite information while rectifying any potential misconceptions or inaccuracies.

Our assessment criteria are stringent and exhaustive. A model is awarded 2 points if it furnishes an answer that is both entirely accurate and inclusive of all pertinent information. A score of 1 point is allotted if the model's response, though accurate, lacks key details. No points are allocated for answers containing factual errors. This scoring framework is crafted to incentivize models to furnish comprehensive, accurate, and nuanced responses while adeptly discerning and rectifying misinformation.

Through this evaluative framework, we not only comprehensively gauge the performance of our large AI teaching assistant model but also pinpoint its strengths and weaknesses in understanding and processing information within specific domains, gleaned from its performance in answering questions. These insights and feedback are invaluable, serving as the bedrock for refining the model and enhancing the efficacy and impact of the AI teaching assistant in subsequent iterations.

Ultimately, by aggregating and analyzing the scores derived from the model's responses to these 80 questions, we will generate a comprehensive evaluation report. This report not only offers a holistic appraisal of the model's performance within a given field but also furnishes guidance for further optimization and enhancement. In so doing, we are steadfast in our commitment to continually augmenting the capabilities of AI teaching assistants, positioning them as indispensable allies in the realms of academic research and education.

### 6.3 Results

For the AI teaching assistant system, evaluating its performance on general tasks is crucial. By conducting experiments on common benchmarks, we can assess WisdoMentor's effectiveness across various task types and problems, thereby determining its utility and reliability in diverse scenarios.

To delve deeper into WisdoMentor's potential in education, a specialized benchmark test called WisdoMentorEval was developed. This test allows for a more precise evaluation of WisdoMentor's performance in educational tasks, such as answering student queries and offering instructional suggestions.In designing a comparative experiment, two groups of comparison models were selected: large models and small models. By comparing with large models (e.g., GPT-3.5, ERNIE Bot-3.5, Kimi), we can gauge WisdoMentor's performance against more robust models, shedding light on its capabilities in handling complex tasks. Conversely, by comparing with small models (e.g., miniCPM, gemma-2B, Phi-2(2B), Qwen-1.8B), we can assess WisdoMentor's efficiency in resource utilization and its adaptability to low-resource environments.

The performance of WisdoMentor on common benchmark is illustrated in Table1 , demonstrating its excellent performance on benchmark datasets. Moreover, drawing from the data presented in Table2 and Table3, we can deduce that WisdoMentor excels in the AI mentor benchmark. These findings further corroborate WisdoMentor's exceptional performance in both general and educational tasks, underscoring its significant potential in the realm of AI teaching assistants.

Table 2: Benchmark performance that specifically evaluates the level of AI teaching assistants in large models

| | Score | ACC | Com | Wrong |
|---|---|---|---|---|
| ChatGPT-3.5 | Score | ACC | Com | Wrong |
| ERNIE Bot-3.5 | Score | ACC | Com | Wrong |
| Kimi | Score | ACC | Com | Wrong |
| WisdoMentor | Score | ACC | Com | Wrong |

## 7 Discussion

In this report, we introduce the WisdoMentor large language model, showcasing the latest advancements in natural language processing. The model has undergone pre-training on massive datasets, encompassing trillions of tokens, and has been fine-tuned utilizing state-of-the-art technologies such

Table 3: Benchmark performance that specifically evaluates the level of AI teaching assistants in small models

| | Score | ACC | Com | Wrong |
|---|---|---|---|---|
| MiniCPM-2B | Score | ACC | Com | Wrong |
| Gemma-2B | Score | ACC | Com | Wrong |
| Phi-2(2B) | Score | ACC | Com | Wrong |
| Qwen-1.8B | Score | ACC | Com | Wrong |
| WisdoMentor | Score | ACC | Com | Wrong |

as SFT, DPO, MoE, and others. Moreover, WisdoMentor incorporates incremental training based on data from academic papers to excel in the teaching assistant domain. Our findings demonstrate that the WisdoMentor family competes favorably with existing open-source models and even rivals the performance of certain proprietary models on synthetic benchmarks and human evaluations. We are confident that making WisdoMentor openly accessible will stimulate collaboration and innovation within the community, enabling researchers and developers to leverage our work and push the boundaries of language models. By democratizing access to these models, we aim to catalyze new research and applications that will significantly propel the field forward and deepen our understanding of the variables and technologies introduced in real-world contexts. In essence, WisdoMentor signifies a significant milestone in the advancement of large-scale language models, and we eagerly anticipate its role in driving progress and innovation in the teaching assistant domain in the years to come.

## Acknowledgments

Thanks for Everyone in WisdoMentor.

## References

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Jiaxi Cui, Zongjian Li, Yang Yan, Bohua Chen, and Li Yuan. 2023. Chatlaw: Open-source legal large language model with integrated external knowledge bases. *arXiv preprint arXiv:2306.16092*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Shuangrui Ding, Zihan Liu, Xiaoyi Dong, Pan Zhang, Rui Qian, Conghui He, Dahua Lin, and Jiaqi Wang. 2024. Songcomposer: A large language model for lyric and melody composition in song generation. *arXiv preprint arXiv:2402.17645*.

Kshitij Gupta, Benjamin Thérien, Adam Ibrahim, Mats L Richter, Quentin Anthony, Eugene Belilovsky, Irina Rish, and Timothée Lesort. 2023. Continual pre-training of large language models: How to (re) warm your model? *arXiv preprint arXiv:2308.04014*.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. *arXiv preprint arXiv:2004.10964*.

Xisen Jin, Dejiao Zhang, Henghui Zhu, Wei Xiao, Shang-Wen Li, Xiaokai Wei, Andrew Arnold, and Xiang Ren. 2021. Lifelong pretraining: Continually adapting language models to emerging corpora. *arXiv preprint arXiv:2110.08534*.

Dahyun Kim, Yungi Kim, Wonho Song, Hyeonwoo Kim, Yunsu Kim, Sanghoon Kim, and Chanjun Park. 2024. sdpo: Don't use your data all at once. *arXiv preprint arXiv:2403.19270*.

OpenAI. 2022. Chatgpt. Accessed: 2024.

Haolin Pan, Yong Guo, Qinyi Deng, Haomin Yang, Jian Chen, and Yiqun Chen. 2023. Improving fine-tuning of self-supervised models with contrastive initialization. *Neural Networks*, 159:198–207.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D Manning, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. arxiv 2023. *arXiv preprint arXiv:2305.18290*.

Luca Soldaini, Rodney Kinney, Akshita Bhagia, Dustin Schwenk, David Atkinson, Russell Authur, Ben Bogin, Khyathi Chandu, Jennifer Dumas, Yanai Elazar, et al. 2024. Dolma: An open corpus of three trillion tokens for language model pretraining research. *arXiv preprint arXiv:2402.00159*.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2022. Self-instruct: Aligning language models with self-generated instructions. *arXiv preprint arXiv:2212.10560*.

Chengyue Wu, Yukang Gan, Yixiao Ge, Zeyu Lu, Jiahao Wang, Ye Feng, Ping Luo, and Ying Shan. 2024. Llama pro: Progressive llama with block expansion. *arXiv preprint arXiv:2401.02415*.

Xuanyu Zhang and Qing Yang. 2023. Xuanyuan 2.0: A large chinese financial chat model with hundreds of billions parameters. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 4435–4439.