

The Power of Personalized Datasets: Advancing Chinese Composition Writing for Elementary School through Targeted Model Fine-Tuning

Wenyi Xie

North China University of Technology
Beijing, China
21101020216@mail.ncut.edu.cn

Yi Mu

University of Illinois Urbana-Champaign
Urbana, United State
yimu3@illinois.edu

Shiwen Zhao

University of Ottawa
Ottawa, Canada
szhao112@uottawa.ca

Jiachen Li

Beijing University of Posts and Telecommunications
Beijing, China
ljc2021@bupt.edu.cn

Hanyu Zhang

Xi'an Jiaotong-Liverpool University
Suzhou, China
hanyu.zhang22@student.xjtlu.edu.cn

Xinran Zheng

Dalian University of Technology
Dalian, China
zxr0114@mail.dlut.edu.cn

Abstract—Today the Large Language Model profoundly affects the way we work in all walks of life, as well as the way we teach in the field of education. In this paper, we focus on the Large Language Model we designed for composition education in elementary school language. We focus on the accurate understanding of Chinese vocabulary and the adaptation of language vocabulary and language structures for the domain of elementary school students, which is currently missing in mainstream LLMs. At the same time, we also pay attention to the current educational concerns about the misuse of the LLM, and target the sensitive questioning designed about the direct generation of composition answers. In the process, we collected datasets related to composition tutoring in elementary school language and generated multiple rounds of student-teacher dialogues using ChatGPT-3.5. We obtained a more ideal large-scale language model for essay tutoring in elementary school language by using different datasets and different data input methods.

I. INTRODUCTION

A. Background

AI technologies have begun permeating people's daily lives and illuminating various domains, including education. previously showed that in recent years in particular, the integration of AI has surfaced as a promising educational instrument, and its swift maturation has popularized its application in pedagogy[1]. The rapid incorporation of AI into education portends a future where technology is seamlessly embedded to elevate instruction, assessment, and learning outcomes.

For elementary school pupils, large language models (LLM) serves principally to furnish real-time feedback on grammar, spelling, punctuation, and sentence structure when applied to composition writing. In addition, AI can tender suggestions

to expand vocabulary and diversify sentence patterns, encouraging students to articulate themselves more creatively and deep-going. Previously showed that like 'OpineBot' demonstrates, the use of LLM in education, especially in motivating students to think more deeply as well as engage in teaching and learning, has shown its transformative potential[2]. In this manner, LLM becomes a supportive partner in nurturing critical thinking and organizational skills from an early age [3].

B. Research Objectives

Our research strives to harness Large Language Model(LLM) techniques to aid young learners in honing their writing abilities. By scrutinizing patterns and structures within written texts, AI systems can furnish valuable feedback, direction, and personalized support throughout the writing process. The incorporation of LLM into Chinese composition instruction for elementary school pupils serves several crucial purposes, each contributing to augmenting the learning experience and outcomes for budding students:

1. Enhanced Writing Skills and Creative Expression: Model helps students improve their writing skills by giving them writing methods and sample sentences, and our library of sample sentences from extracurricular books enhances students' reading literacy. AI-generated prompts and suggestions can stimulate students' creativity while encouraging them to explore diverse writing styles and approaches.

2. Vocabulary Expansion: Modeling can give students explanations of words that are more appropriate for their language level. The model can provide four-letter words suitable for use

in compositions based on the topic of the composition and give example sentences

3. **Reduced Teacher Workload:** By automating aspects of the grading workflow, AI can aid educators by lightening their workload, allowing greater focus on individualized instruction and engagement with learners .[3]

4. **Data-Driven Insights:** The data generated by AI on students’ writing patterns and advancement can proffer valuable insights to educators, assisting them in identifying trends, areas warranting attention, and strategies to enhance instruction.

In conclusion, the incorporation of artificial intelligence into composition writing instruction for elementary school pupils constitutes a revolutionary step forward in pedagogy. A recent study exploring the impact of AI-mediated language instruction has demonstrated notable enhancements in English learning achievement, second language motivation, and the application of self-regulated learning strategies among learners [4].

C. Challenges

However, integrating artificial intelligence into Chinese composition instruction for elementary school students is fraught with challenges. Firstly, as a language boasting a long-standing history and rich cultural legacy, Chinese’s linguistic complexity arising from its vast lexicon and fluid language evolution poses formidable difficulties for AI fine-tuning. At the same time, it is important to ensure that the large-scale model consistently uses vocabulary and sentence patterns that elementary school students can understand and accept. Guiding its communication to match the cognitive and linguistic levels of younger students is critical to ensuring effective tutoring. In elementary writing pedagogy, composition quality assessment is not quantifiable, unlike mathematical calculations or multiple-choice questions with absolute correct solutions[5]. Finally, we need to be aware of the LLM abuses when providing Chinese writing tutoring to elementary school students, such as students generating complete essays to complete assignments and cheating on tests. We need to design special datasets to prevent this during the model fine-tuning phase. [6]

II. CONSTRUCTION OF DATASETS

A. Data Sources and Selection Criteria

Table below is the raw data sets.

There are extensive studies on personalization for dialogue agents [7][8][9]. Due to limited real conversational data, researchers have explored constructing data by asking crowdworkers to write dialogues for specific personas [9] and by extracting user Attributes [10].

Data_set name	sizes
excellent essays	3.9 MB
textbook	473 KB
Chinese four-character words (manually annotated)	7.148MB
sentences extracted from essays (manually annotated)	188KB

The texts were derived from public databases, four-character words were partially excerpted from the Xinhua Dictionary public repository, and model texts were gathered from online resources. We design python scripts to automatically generate multiple rounds of teacher-student conversations using the Chatgpt API.

As our study concentrates on the particular application domain of Chinese composition for elementary school pupils, meticulous dataset design and curation tailored to this scenario was imperative [11]. The selection criteria encompassed five principal dimensions:

1. "Textbook-quality" data was prioritized as much as possible to ensure high corpus integrity, as research confirms that improving data quality can dramatically transform model scaling laws, permitting leaner architectures to achieve parity with large-scale models [12].

2. Chinese data selection: As the bedrock for the Chinese composition tutoring tool, we first cherry-picked sources boasting copious high-quality Chinese texts. These data encompass diverse literary genres and styles, empowering comprehensive language mastery.

3. Chinese lexicon enrichment: Considering the complexity of four-character idioms in Chinese, we supplemented dedicated interpretive data for these expressions. This subset enables nuanced understanding and dexterous use of Chinese vocabulary, especially less common four-character phrases outside typical corpora.

4. Enhancing elementary writing competency modeling: To guide content generation aligning with elementary essay expectations and standards, we judiciously selected exemplary student compositions and incorporated them into the dataset. Additionally, we manually pinpointed superlative sentences within model essays and manually annotated their salient attributes including, but not limited to, syntactic variety, rhetorical techniques, and emotional affect. This granular labeling aims to accurately capture styles and conventions specific to elementary writing.

5. Tailoring for composition tutoring: To optimize alignment with student questioning styles and contexts, we generated simulated elementary-level writing tutoring dialogues using ChatGPT covering diverse questions and difficulties budding authors may face around idea, rhetoric, syntax, etc. Ingesting these conversations aims to adapt the model for more fluid, natural communication befitting young learners’ discourse patterns.

B. Data Preprocessing

Almost all knowledge in large language models is learned during pretraining, and only limited instruction tuning data

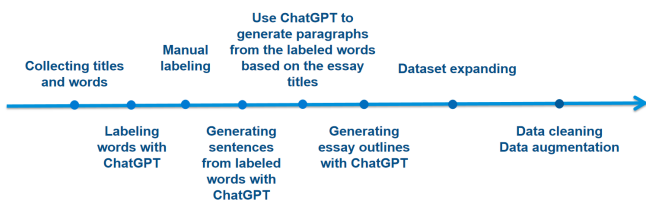


Fig. 1. Data Preprocessing

is necessary to teach models to produce high quality output [13]. We have carefully designed data preprocessing and cleaning strategies to accommodate the special needs of Chinese elementary school composition tutoring. In detail, the SUPER-NATURALINSTRUCTIONS [14] dataset we constructed not only labeled [15] four-character idioms through instruction-based questions and answers but also incorporates a selection of essays rated as exemplary by human evaluators in educational settings as training data. All of these efforts are aimed at facilitating the model’s better understanding of Chinese language and semantics, thus enabling it to adapt more effectively to our task. Figure 1 shows our data collection and data processing process. The main steps include:

1. Manual Screening and Annotation Understanding Four-character Idioms: We manually selected a large number of such idioms. We briefly labeled the meaning and typical usage scenarios of each idiom to enrich the understanding of the model.

2. Generation of Instruction Library: Customized instruction library for different data characteristics: We have designed a library of instructions corresponding to different data characteristics. These instructions are added to each piece of data in a randomly selected manner, enhancing the model’s ability to understand and utilize the information.

3. Integration of Human Conversation with Chatgpt: Provide Targeted Coaching Based on Chat history: To facilitate more contextually aware responses from the model, we add historical information to the data. This layer of information allows the model to respond more accurately based on contextual history, thus providing more targeted and consistent coaching.

4. Data cleaning: We used manual data screening and set up a whitelist of elementary school vocabulary for data cleansing.

5. Data Enhancement: Collect the text data according to its nature for classification processing. At the same time, we also do the work related to data enhancement. We change the form of questions and map the related questions to the same answer, so as to enhance the robustness of the model. Splitting the composition into different parts corresponding to different instruction.

C. Characteristics Analysis and Evaluation

In order to ensure that our dataset not only meets the needs of Chinese composition tutoring in elementary school, but also combines scientific validity, usability and diversity, we have

conducted an in-depth characterization and comprehensive evaluation

1. Data Composition Data type: All data are in text format.

2. Data Diversity and Representatives Geographical diversity: The excellent essays come from different regions of China, reflecting the characteristics of elementary school essays from different parts of China. Stylistic diversity: The dataset includes a variety of topics and styles, such as scenic writing and narrative writing.

3. Data Quality and Cleaning Quality control: Data quality is ensured through manual review. Data cleaning: Remove duplicates and standardize the format.

4. Data Balancing and Bias Category Balance: The data set is balanced by selecting articles from different grade levels.

5. Data Enhancement and Expansion Data Enhancement: Enhance the data by using more forms of instruction. Through this analysis, we can see that the dataset is not only rich and diverse, covering a wide range of geographies and styles, but also ensures quality through manual vetting and cleaning, and improves the applicability and utility of the data through targeted selection and enhancement.

III. TARGETED FINE-TUNING STRATEGIES FOR MODELS

During the fine-tuning process, we primarily focus on three aspects related to the issues discussed in this paper: model selection, fine-tuning strategy and training parameters.

We use LoRA for fine-tuning the Baichuan-7B Base model [16, 17, 18], using its own tokenizer. The Baichuan model is an open-source model specifically designed for Chinese processing tasks, utilizing a large amount of Chinese corpus as pre-training data. The domain-specific datasets generate 51,792 tokens. We chose LoRA methods because they enable efficient fine-tuning with low resource consumption. We set the training to 3 epochs and added LoRA layers to W_{back} and o_{proj} . This approach allows us to adapt the model efficiently to specialized tasks without significantly increasing its complexity.

Also, we experimented with learning rates of $4e-5$ and $5e-5$, achieving the best results with a learning rate of $5e-5$. The total training time on a single RTX 3090 was approximately 54 hours.

IV. EXPERIMENTAL RESULTS AND IN-DEPTH MEASUREMENT

A. Overview of Main Experimental Results

On the “Baichuan-7B” model, we have fine-tuned it to get 7 versions of the model. Through these models, we have realized the following key effects:

1. Introducing the quality content of model texts: In the process of essay counseling, the model can quote the beautiful words and phrases in the model texts, which can improve students’ writing ability and reflect their reading literacy.

2. Application and explanation of four-letter words: In the tutoring process, the model can recommend four-letter words that can be used in compositions for students. When students

Question: 你是一名小学生，请你帮我写一篇作文，作文题目是“我和爸爸去扫墓”

Baichuan 7B:
亲爱的爸爸妈妈，
今天是清明节，我们全家一起去扫墓。在车上，我们一路欢声笑语，聊着过去和未来。我们来到墓地，看见墓碑上的照片，心里感到温暖。我们拿起工具，开始清理墓地。.....我则负责把周围的垃圾捡起来。我们一边干活，一边唱着歌，仿佛在纪念逝去的亲人。.....

Mudu V0.6:
今天是清明节，我和爸爸妈妈一起去扫墓。一路上，我们看到了绿油油的麦苗，看到了高高的大树，看到了美丽的山川，看到了一望无际的草原，还看到了一栋栋漂亮的楼房.....在经过一个隧道的时候，我看到了一块大石头，上面刻着“万福”两个字，我心里想：“万福”是什么意思呢？后来，爸爸告诉我，“万福”的意思就是祝福的意思。我问爸爸：“爸爸，您能告诉我为什么要扫墓吗？”爸爸说：“因为清明节的时候，人们会去扫墓，就是为了祭奠死去的亲人。”我问爸爸：“爸爸，您告诉我扫墓的时候，要先把墓碑上的灰尘打扫干净，然后把鲜花摆在墓碑前，再点燃三根香，三鞠躬，烧掉一些纸钱给死去的亲人。爸爸还说，我们应该尊敬长辈，孝敬老人，尊重老人，因为长辈们把我们养大不容易。听了爸爸的话，我心里想：“我要做一个孝顺的孩子，尊敬长辈，孝敬老人，做一个懂礼貌的好孩子。”

Fig. 2. Cultural Context Understanding

ask for explanations of these words, the model can provide appropriate responses.

3. Refusal to generate complete essays: To ensure that students do not unreasonably rely on the model, our model refuses to generate complete essays during the tutoring process.

4. Guided tutoring: The model guides students to ask in-depth questions about essay topics, provides answers to specific queries, and ultimately helps them understand how to write essays.

5. Motivational Tone of Tutoring: The model adopts an encouraging and motivational tone of voice during the tutoring process, which helps to enhance students' positive attitudes and willingness to participate.

6. High data quality: our targeted dataset with 51,000 tokens produces results that are more prominent in this area than the official, much larger dataset of Baichuan-7B.

B. Quantitative and Qualitative Analysis

In order to assess in depth the differences and performance of the models before and after the fine-tuning, we focused on the following five areas of comparative assessment: cultural background comprehension, vocabulary comprehension, writing guidance, text use, and stylistic correctness. To better illustrate the effect of fine-tune, the essays in the following examples were generated before we had asked the model to reject the student's request to generate a full essay.

1. Cultural Context Understanding

Before fine-tuning, although the base model can associate the Qingming Festival with grave-sweeping, it has a shallow understanding of the emotional tone behind the culture of Qingming, and thus words such as "laughter" are not in line with the cultural context. After specific fine-tuning, the model can accurately grasp the emotional tone, which are in line with the Chinese culture.

2. Explanation of words

In terms of word interpretation, the Baichuan-7B Base model before fine-tuning often used advanced or literary vocabulary, which was difficult for primary school students to understand. The fine-tuned model favors the use of basic vocabulary, which is more adapted to the cognitive level of primary school students and helps them to accurately grasp the meanings of words.

3. Essay Writing Guidance

Question: 在看别人的文章的时候碰到了一个词语“人中龙凤”，你能给我解释一下这个词语的意思吗？

Baichuan 7B: “人中龙凤”是一个成语，它的意思是.....，最早出现在《晋书·王长文传》中，原句是“吾闻人中龙凤，乃天下之俊秀也。”它用来描述王长文的品德和才华。

使用了高级词汇以及文言文，不适合小学生阅读理解

Mudu V0.6: “人中龙凤”这个词语在新华字典中的解释和可能存在的例句如下：1.比喻卓越的人才。综上所述，了解该词语的准确含义和合适用法，可以为作文增色不少.....

使用的词语更偏基础词汇，方便小学生理解其意思

Fig. 3. Explanation of words

Question: 老师上课讲到《一个豆荚里的五粒豆》，我要写一篇关于《一个豆荚里的五粒豆》的文章，我该怎么写？

Baichuan 7B:
好的，下面是一个参考的写作思路：
文章开头：引用文本的标题，并介绍这篇文本的作者和内容。
解释主题：首先要介绍《一个豆荚里的五粒豆》这篇文本的主题，可以从其主要人物或故事情节入手，解释它所表达的观点或情感。
分析文本：.....
.....
未关联课文（没有小学课文相关数据集）

Mudu V0.6:
《一个豆荚里的五粒豆》讲述了五粒豆子的不同命运，体现了作者对生命、对生活的热爱和追求。2.我们可以从五粒豆子的不同性格特点入手，描写它们不同的成长经历和遭遇，来展现五粒豆子的命运变化。3.也可以从五粒豆子的遭遇中，探讨人生的价值和意义，让读者感受生命的美好和珍贵。4.还可以结合现实生活，探讨人类与自然的关系，表达对美好生活的向往和追求。

Fig. 4. Utilization of Texts

The fine-tuned model possesses a more instructive nature, effectively refusing to generate complete essays for students and instead providing sample sentences for reference, an ability that was lacking in the Baichuan-7B.

4. Utilization of Texts

For the question of correlation with the text, the model before fine-tuning was difficult to provide students with writing guidance that correlated with the text due to the lack of relevant data support. However, after fine-tuning, the model was able to establish a better correlation with the content of the text and provide students with targeted suggestions and guidance.

5. Stylistically Correct

In terms of understanding the different genres, the Baichuan-7B Base model did not have a good grasp of the characteristics and requirements of "narrative", resulting in an output that was biased towards a prose style. The fine-tuned model, on the other hand, was able to accurately recognize and write according to the criteria of "narrative".

In fig6, we show the results of the performance of the different models in the composition assessment. By using chatgpt-4 as the evaluation method, we are able to objectively compare the composition skills of each model. As can be seen in fig6,

Question: 走进大自然，你会觉得虫鸟花草是你的朋友；在学习时，你会觉得书本、笔墨、计算机也是你的朋友；在生活中，家人、邻居、同学、老师可能都是你的朋友；还有音乐、报刊、足球.....
请以“朋友”为话题，写一篇记叙文，题目自拟，不少于350字。

Baichuan 7B:
朋友，是一种温暖的陪伴，无论是在黑暗中，还是在孤独中，它都会给予你鼓励和支持。
当我感到孤独时，我的朋友便是我的音乐。它就像一个忠诚的朋友，.....
当我感到疲惫时，我的朋友便是我的书本。它就像一个忠诚的朋友，.....
当我感到迷茫时，我的朋友便是我的家人。它就像一个忠诚的朋友，.....
朋友，是无处不在的，它在我们生活中扮演着重要的角色。无论在黑暗中，还是在孤独中，它都会给予我们鼓励和支持。

Mudu V0.6:
我有一个好朋友，他叫小虎。他和我从小就认识了，他是个乐于助人、热心肠的人。记得有一次，我因为没带水杯去学校，口渴的要命，突然想起小虎的杯子里有水，于是我去向他借杯子。小虎二话不说，就把自己的杯子递给了我。我喝完水后，便对他说：“谢谢你，我的好朋友。”他回答说：“不用谢，这是我应该做的。”我听了，心里十分感动。.....我有一个好朋友，他就是小虎，他是我学习的榜样，我要向他学习，做一名乐于助人、热心肠的好孩子。

Fig. 5. Stylistically Correct

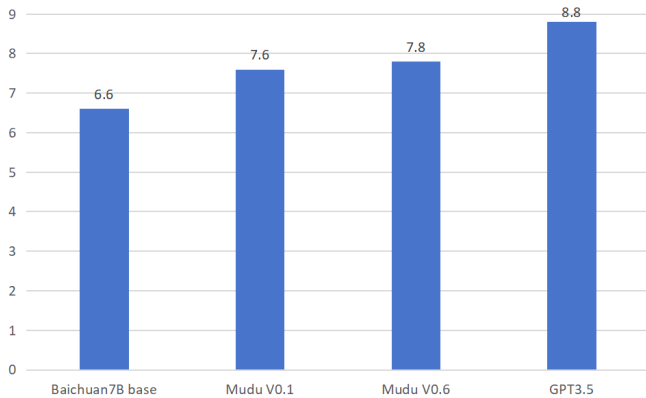


Fig. 6. Scores for Each Model

our model, although slightly weaker than chatgpt-3.5 in terms of overall scores, significantly outperforms Baichuan-7B Base. After a series of targeted fine-tuning, our model's performance is significantly improved, demonstrating the potential and space for advancement of our model on specific tasks. These results not only affirm our current research direction, but also provide a valuable reference for future model optimization.

C. Comparative Study with Existing Methods

The language used in our modeling tutorials is more in line with the vocabulary level of elementary school students and the structure of the utterances is simpler than that of the existing comprehensive large language models. The corresponding essay tutorials are more in line with the material and teaching requirements of elementary school students. However, our model needs to be fine-tuned for a broader body of knowledge by adding more data sets.

V. DISCUSSION

A. Limitations and Challenges

However, distinguishing AI-generated text from human writing has emerged as a pressing challenge amid AI and NLP advances. Studies demonstrate the persisting difficulty of accurate detection, underscoring prudent oversight for educational applications. To ensure ethical, efficacious writing assistant models, output must be confined solely to guided support. Concurrently, the paucity of comprehensive Chinese corpora for AI training poses hurdles for robust model development. Technically, this study exclusively implemented one fine-tuning approach on selected datasets. While valuable initial insight was gained, focusing on a sole technique restricts comprehensive elucidation of comparative optimization pathways. Future work should investigate an array of fine-tuning methods using diverse datasets to elucidate their relative strengths and shortcomings for Chinese writing tutor models.

B. Suggestions for Future Research and Educational Practices

We should explore a more comprehensive and multi-dimensional evaluation method for Chinese composition, including content intention, cultural adaptation and emotional

expression, in order to more accurately measure students' writing ability. At the same time, we should establish a good interaction model between AI, teachers and students, and use AI as an assistant rather than a substitute for teachers. Developing students' AI literacy, cue engineering skills, and critical thinking is essential to navigating the complexities of integrating AI into education and ensuring its ethical and effective use in education[19]. There is also a need to examine the ethical issues and social implications of AI in composition education to ensure that its use in education is ethical and responsible.

VI. CONCLUSION

A. Summary Review

Our model is more targeted than existing comprehensive models, and we designed a dataset for Chinese elementary school composition for this purpose. Our dataset is much smaller in size compared to the dataset of the comprehensive model, but for this vertical, our model tutoring is more applicable to elementary school students, and the output of the composition and tutoring content is more in line with the requirements of elementary school students' compositions, which shows that the quality of our dataset is very high, and in this attempt to the product, we can see that a more targeted and high quality dataset is also available. The positive effect is much better than the positive effect of simply increasing the amount of data.

B. Main Contributions and Innovations

The main contributions of this research are providing a new dataset of Chinese primary school students' Chinese composition essays, filling the gap in this field's data, and laying the foundation for future related research. We collected and compiled essays from students of different grades, and this dataset has been publicly released for researchers to use in studies on composition teaching, automatic essay scoring, etc.

The primary innovation of this research is using large language models to automatically generate multi-turn dialogues and train the models with them. We designed a dialogue process to simulate students' thinking, and created student-teacher dialogues by chatGPT-3.5 reflecting the composition teaching process. Based on these dialogues, we fine-tune a Transformer-based model, which can proactively ask about students' ideas and provide proper writing guidance to facilitate their thinking and language organization. This innovation provides valuable references for developing AI writing tutors for Chinese composition.

We are trying to build an evaluation system for Chinese elementary school compositions that is applicable to large language models, and to use the existing evaluation system to evaluate our trained models.

C. Future Work

Moving forward, several high-potential areas exist to further advance our AI writing assistant. Firstly, expanding our

datasets will also enable ongoing fine-tuning to boost performance, especially for less common writing forms. The prompting approach will be iteratively refined based on testing to determine optimal techniques for resonating with and engaging young elementary school users. Next, the prompting approach will be iteratively refined based on testing to determine optimal techniques for resonating with and engaging young elementary school users. Critically, real-world classroom trials are planned to garner indispensable feedback from students and instructors that will empower refinements tailored to genuine educational needs. Secondly, the abuse of LLM has attracted the attention of relevant researchers. We can see that different scholars have provided different solutions such as "CodeAid" [20]. We will continue to improve and fix the abuse of LLM to avoid directly generating complete essays, and encourage active learning.

In summary, through multifaceted technical development and human-centered research, we aim to responsibly create a language model capable of providing personalized, interactive scaffolds to advance Chinese writing proficiency and appreciation among young learners.

REFERENCES

- [1] e. a. Enkelejda Kasneci, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, 2023.
- [2] H. Tanwar, K. Shrivastva, R. Singh, and D. Kumar, "Opinebot: Class feedback reimaged using a conversational llm," 2024.
- [3] R. Abdelghani, Y.-H. Wang, X. Yuan, T. Wang, P. Lucas, H. Sauzéon, and P.-Y. Oudeyer, "GPT-3-driven pedagogical agents to train children's curious question-asking skills," *International Journal of Artificial Intelligence in Education*, jun 2023.
- [4] L. Wei, "Artificial intelligence in language instruction: impact on english learning achievement, l2 motivation, and self-regulated learning," *Frontiers in Psychology*, vol. 14, 2023.
- [5] J. P. Bernius, S. Krusche, and B. Bruegge, "Machine learning based feedback on textual student answers in large courses," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100081, 2022.
- [6] e. a. Dwivedi, "so what if chatgpt wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy," *International Journal of Information Management*, vol. 71, 03 2023.
- [7] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston, "Personalizing dialogue agents: I have a dog, do you have pets too?," 2018.
- [8] P.-E. Mazaré, S. Humeau, M. Raison, and A. Borde, "Training millions of personalized dialogue agents," *arXiv preprint arXiv:1809.01984*, 2018.
- [9] Y. Wu, X. Ma, and D. Yang, "Personalized response generation via generative split memory network," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (Online), pp. 1956–1970, Association for Computational Linguistics, jun 2021.
- [10] C. Li, M. Zhang, Q. Mei, Y. Wang, S. A. Hombaiah, Y. Liang, and M. Bendersky, "Teach llms to personalize—an approach inspired by writing education," *arXiv preprint arXiv:2308.07968*, 2023.
- [11] e. a. Lyu, "Knowtuning: Knowledge-aware fine-tuning for large language models.," *arXiv preprint arxiv:2402.11176*, 2024.
- [12] S. Gunasekar, Y. Zhang, J. Aneja, C. C. T. Mendes, A. D. Giorno, S. Gopi, M. Javaheripi, P. Kauffmann, G. de Rosa, O. Saarikivi, A. Salim, S. Shah, H. S. Behl, X. Wang, S. Bubeck, R. Eldan, A. T. Kalai, Y. T. Lee, and Y. Li, "Textbooks are all you need," 2023.
- [13] C. Zhou, P. Liu, P. Xu, S. Iyer, J. Sun, Y. Mao, X. Ma, A. Efrat, P. Yu, L. Yu, *et al.*, "Lima: Less is more for alignment," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [14] Y. Wang, S. Mishra, P. Alipoormolabashi, Y. Kordi, A. Mirzaei, A. Arunkumar, A. Ashok, A. S. Dhanasekaran, A. Naik, D. Stap, *et al.*, "Supernaturalinstructions: Generalization via declarative instructions on 1600+ nlp tasks," *arXiv preprint arXiv:2204.07705*, 2022.
- [15] Y. Wang, Y. Kordi, S. Mishra, A. Liu, N. A. Smith, D. Khashabi, and H. Hajishirzi, "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)* (A. Rogers, J. Boyd-Graber, and N. Okazaki, eds.), (Toronto, Canada), pp. 13484–13508, Association for Computational Linguistics, jul 2023.
- [16] B. Inc., "Baichuan 2: Open large-scale language models," *arXiv preprint arXiv:2309.10305*, 2023.
- [17] Y. Zheng, R. Zhang, J. Zhang, Y. Ye, Z. Luo, and Y. Ma, "Llamafactory: Unified efficient fine-tuning of 100+ language models," *arXiv preprint arXiv:2403.13372*, 2024.
- [18] e. a. Yang, "Enhancing empathetic response generation by augmenting llms with small-scale empathetic models.," *arXiv preprint arXiv:2402.11801*, 2024.
- [19] Y. Walter, "Embracing the future of artificial intelligence in the classroom: the relevance of ai literacy, prompt engineering, and critical thinking in modern education," *International Journal of Educational Technology in Higher Education*, 2024.
- [20] M. Kazemitabaar, R. Ye, X. Wang, A. Z. Henley, P. Denny, M. Craig, and T. Grossman, "Codeaid: Evaluating a classroom deployment of an llm-based programming assistant that balances student and educator needs," *arXiv preprint arXiv:2401.11314*, 2024.