refinement (Liu et al., 2023c), and code question answering (Liu & Wan, 2021). In a word, LLMs have the potential to revolutionize the field of coding by providing developers with powerful tools for code comprehension, generation, and related tasks.

6.5 LLM FOR MATHEMATICS

LLMs with a certain model scale have been found to possess the ability to perform mathematical reasoning (Wei et al., 2022b; Suzgun et al., 2022). In order to encourage LLMs to achieve better performance on math-related tasks, researchers have employed techniques such as chain-of-thought prompting (Wei et al., 2022c) and scratchpad (Nye et al., 2021), which have shown promising results. Additionally, self-consistency (Wang et al., 2022) and least-to-most prompting (Zhou et al., 2022) have further improved the performance of these models on these tasks. However, prompt engineering is a time-consuming process that requires a lot of trial and error, and it is still difficult for LLMs to consistently perform well or achieve satisfactory results in solving mathematical problems. Moreover, simply scaling the data and model size is not an efficient way to improve a model's mathematical reasoning abilities. Instead, pretraining on math-related corpora has been shown to consistently enhance these capabilities (Hendrycks et al., 2021; Lewkowycz et al., 2022; Taylor et al., 2022; Lightman et al., 2023). Additionally, fine-tuning on math-related instruction-following datasets (Si et al., 2023; Yuan et al., 2023a; Luo et al., 2023a; Yue et al., 2023; Chern et al., 2023a; Yu et al., 2023), has also been effective and more cost-effective than math-specific pretraining. Despite their limitations in terms of accuracy, LLMs still have significant potential to assist users with practical mathematical problems. There is ample scope for further development in this area.

7 CONCLUSION

In this report, we present the QWEN series of large language models, which showcase the latest advancements in natural language processing. With 14B, 7B, and 1.8B parameters, these models have been pre-trained on massive amounts of data, including trillions of tokens, and fine-tuned using cutting-edge techniques such as SFT and RLHF. Additionally, the QWEN series includes specialized models for coding and mathematics, such as CODE-QWEN, CODE-QWEN-CHAT, and MATH-QWEN-CHAT, which have been trained on domain-specific data to excel in their respective fields. Our results demonstrate that the QWEN series is competitive with existing open-source models and even matches the performance of some proprietary models on comprehensive benchmarks and human evaluation.

We believe that the open access of QWEN will foster collaboration and innovation within the community, enabling researchers and developers to build upon our work and push the boundaries of what is possible with language models. By providing these models to the public, we hope to inspire new research and applications that will further advance the field and contribute to our understanding of the variables and techniques introduced in realistic settings. In a nutshell, the QWEN series represents a major milestone in our development of large language models, and we are excited to see how it will be used to drive progress and innovation in the years to come.