QWEN TECHNICAL REPORT

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, Jianxin Ma, Rui Men, Xingzhang Ren, Xuancheng Ren, Chuanqi Tan, Sinan Tan, Jianhong Tu, Peng Wang, Shijie Wang, Wei Wang, Shengguang Wu, Benfeng Xu, Jin Xu, An Yang, Hao Yang, Jian Yang, Shusheng Yang, Yang Yao, Bowen Yu, Hongyi Yuan, Zheng Yuan, Jianwei Zhang, Xingxuan Zhang, Yichang Zhang, Zhenru Zhang, Chang Zhou, Jingren Zhou, Xiaohuan Zhou, Tianhang Zhu.

Qwen Team, Alibaba Group*

ABSTRACT

Large language models (LLMs) have revolutionized the field of artificial intelligence, enabling natural language processing tasks that were previously thought to be exclusive to humans. In this work, we introduce QWEN[1], the first installment of our large language model series. QWEN is a comprehensive language model series that encompasses distinct models with varying parameter counts. It includes QWEN, the base pretrained language models, and QWEN-CHAT, the chat models finetuned with human alignment techniques. The base language models consistently demonstrate superior performance across a multitude of downstream tasks, and the chat models, particularly those trained using Reinforcement Learning from Human Feedback (RLHF), are highly competitive. The chat models possess advanced tool-use and planning capabilities for creating agent applications, showcasing impressive performance even when compared to bigger models on complex tasks like utilizing a code interpreter. Furthermore, we have developed coding-specialized models, CODE-QWEN and CODE-QWEN-CHAT, as well as mathematics-focused models, MATH-QWEN-CHAT, which are built upon base language models. These models demonstrate significantly improved performance in comparison with open-source models, and slightly fall behind the proprietary models.

*Authors are ordered alphabetically by the last name. Correspondence to: ericzhou.zc@alibaba-inc.com.

[1]QWEN is a moniker of Qianwen, which means "thousands of prompts" in Chinese. The pronunciation of "QWEN" can vary depending on the context and the individual speaking it. Here is one possible way to pronounce it: /kwɛn/.

Contents