

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 3 |
| 2 | Pretraining | 4 |
| 2.1 | Data | 4 |
| 2.2 | Tokenization | 6 |
| 2.3 | Model | 6 |
| 2.3.1 | Architecture | 6 |
| 2.3.2 | Context Length Extension | 7 |
| 2.4 | Training | 8 |
| 2.5 | Experimental Results | 8 |
| 3 | Alignment | 9 |
| 3.1 | Supervised Finetuning | 9 |
| 3.1.1 | Data | 9 |
| 3.1.2 | Training | 10 |
| 3.2 | Reinforcement Learning from Human Feedback | 10 |
| 3.2.1 | Reward Model | 10 |
| 3.2.2 | Reinforcement Learning | 11 |
| 3.3 | Automatic and Human Evaluation of Aligned Models | 11 |
| 3.4 | Tool Use, Code Interpreter, and Agent | 13 |
| 4 | CODE-QWEN: Specialized Model for Coding | 16 |
| 4.1 | Code Pretraining | 16 |
| 4.2 | Code Supervised Fine-Tuning | 17 |
| 4.3 | Evaluation | 17 |
| 5 | MATH-QWEN: Specialized Model for Mathematics Reasoning | 17 |
| 5.1 | Training | 17 |
| 5.2 | Evaluation | 20 |
| 6 | Related Work | 20 |
| 6.1 | Large Language Models | 20 |
| 6.2 | Alignment | 20 |
| 6.3 | Tool Use and Agents | 21 |
| 6.4 | LLM for Coding | 21 |
| 6.5 | LLM for Mathematics | 22 |
| 7 | Conclusion | 22 |
| A | Appendix | 35 |
| A.1 | More Training Details | 35 |
| A.1.1 | Data Format for QWEN-CHAT | 35 |
| A.2 | Evaluation | 35 |
| A.2.1 | Automatic Evaluation | 35 |
| A.2.2 | Human Evaluation | 40 |
| A.3 | Analysis of Code Interpreter | 58 |