

图 1 图片 1

表 1 表格 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 模型名称 | 调用总量 | 调用失败量 | 失败率 | 平均调用时长 | 平均首Token延时 |
| deepseek-r1 | 547 | 115 | 21.02% | 51.28s | 36.01s |
| qwen-long | 62 | 0 | 0.00% | 1.84s | 1.84s |
| qwen-max | 31 | 0 | 0.00% | 15.5s | 4.68s |
| qwen-max-2025-01-25 | 24 | 0 | 0.00% | 29.82s | 3.95s |
| qwen-max-latest | 111 | 0 | 0.00% | 42.96s | 0.57s |
| qwen-plus | 378 | 63 | 16.67% | 3.58s | 3.18s |
| qwen-plus-2025-01-25 | 41 | 0 | 0.00% | 9.94s | 9.94s |
| qwen-turbo | 137 | 0 | 0.00% | 0.75s | 0.75s |
| chatGLM-4-plus | 62 | 0 | 0.00% | 3.31s | 1.11s |
| moonshot-v1-32k | 62 | 0 | 0.00% | 7.33s | 0.99s |
| yi-large | 211 | 1 | 0.47% | 0.96s | 1.67s |

这是第一行：

图表 1这是图1

这是第二行

图表 2 这是表2