

0512fbfd-08d5-462a-8e04-1d7a861d5d94

| dataset     | total_problems | passed_problems | failed_problems | no_test_problems | accuracy_percent |
|-------------|----------------|-----------------|-----------------|------------------|------------------|
| codecontest | 18             | 5               | 13              | 0                | 27.7777777777778 |
| humaneval   | 40             | 27              | 11              | 2                | 67.5             |
| mbpp        | 63             | 50              | 12              | 1                | 79.3650793650794 |

0512fbfd-08d5-462a-8e04-1d7a861d5d94

| total_tests_run | total_tests_passed | total_tests_failed | total_tests_timeout |
|-----------------|--------------------|--------------------|---------------------|
| 3403            | 1285               | 2114               | 4                   |
| 12667           | 8611               | 4056               | 0                   |
| 2231            | 2151               | 80                 | 0                   |