Qwen-2.5-14B-Instruct

Legend: Error Matching, No Error Matching

RR: 94.67% (Error Matching), 73.34% (No Error Matching)
TSR: 79.48% (Error Matching), 66.32% (No Error Matching)
CSR: 94.57% (Error Matching), 68.91% (No Error Matching)