

problem_id	iteration	dataset_type	status	passed_public	failed_public	timeout_public
	0 iteration_0	test	FAILED	0	1	0
	1 iteration_0	test	FAILED	2	0	0
	2 iteration_0	test	FAILED	0	3	0
	3 iteration_0	test	FAILED	1	4	0
	4 iteration_0	test	FAILED	0	2	0
	0 iteration_0	valid	FAILED	0	2	0
	1 iteration_0	valid	FAILED	0	2	0
	2 iteration_0	valid	FAILED	1	2	0
	3 iteration_0	valid	PASSED	4	0	0
	4 iteration_0	valid	PASSED	1	0	0
	5 iteration_0	valid	PASSED	1	0	0
	6 iteration_0	valid	FAILED	0	2	0
	7 iteration_0	valid	FAILED	0	1	0
	8 iteration_0	valid	PASSED	1	0	0
	9 iteration_0	valid	FAILED	0	1	0
	10 iteration_0	valid	FAILED	0	1	0
	11 iteration_0	valid	PASSED	1	0	0
	12 iteration_0	valid	FAILED	0	1	0
	0 iteration_0	plus	FAILED	2	0	0
	1 iteration_0	plus	FAILED	0	1	0
	2 iteration_0	plus	NO_TESTS	0	0	0
	3 iteration_0	plus	FAILED	2	0	0
	4 iteration_0	plus	PASSED	1	0	0
	5 iteration_0	plus	FAILED	1	1	0
	6 iteration_0	plus	FAILED	1	0	0
	7 iteration_0	plus	PASSED	2	0	0
	8 iteration_0	plus	PASSED	2	0	0
	9 iteration_0	plus	FAILED	1	0	0
	10 iteration_0	plus	FAILED	1	2	0
	11 iteration_0	plus	PASSED	1	0	0
	12 iteration_0	plus	FAILED	3	0	0
	13 iteration_0	plus	PASSED	2	0	0
	14 iteration_0	plus	FAILED	1	0	0
	15 iteration_0	plus	PASSED	2	0	0
	16 iteration_0	plus	FAILED	2	0	0
	17 iteration_0	plus	PASSED	1	0	0
	18 iteration_0	plus	PASSED	3	0	0
	19 iteration_0	plus	PASSED	1	0	0
	0 iteration_0	raw	PASSED	2	0	0
	1 iteration_0	raw	NO_TESTS	0	0	0
	2 iteration_0	raw	PASSED	1	0	0
	3 iteration_0	raw	FAILED	2	0	0
	4 iteration_0	raw	PASSED	1	0	0
	5 iteration_0	raw	PASSED	2	0	0
	6 iteration_0	raw	PASSED	1	0	0
	7 iteration_0	raw	PASSED	2	0	0
	8 iteration_0	raw	PASSED	2	0	0
	9 iteration_0	raw	PASSED	1	0	0
	10 iteration_0	raw	PASSED	3	0	0
	11 iteration_0	raw	PASSED	1	0	0
	12 iteration_0	raw	PASSED	3	0	0
	13 iteration_0	raw	PASSED	2	0	0

14 iteration_0	raw	PASSED	1	0	0
15 iteration_0	raw	PASSED	2	0	0
16 iteration_0	raw	PASSED	2	0	0
17 iteration_0	raw	PASSED	1	0	0
18 iteration_0	raw	PASSED	3	0	0
19 iteration_0	raw	PASSED	1	0	0
0 iteration_0	plus	NO_TESTS	0	0	0
1 iteration_0	plus	PASSED	5	0	0
2 iteration_0	plus	PASSED	3	0	0
3 iteration_0	plus	FAILED	6	0	0
4 iteration_0	plus	FAILED	3	0	0
5 iteration_0	plus	PASSED	3	0	0
6 iteration_0	plus	PASSED	3	0	0
7 iteration_0	plus	PASSED	3	0	0
8 iteration_0	plus	PASSED	3	0	0
9 iteration_0	plus	PASSED	3	0	0
10 iteration_0	plus	PASSED	3	0	0
11 iteration_0	plus	PASSED	3	0	0
12 iteration_0	plus	PASSED	3	0	0
13 iteration_0	plus	PASSED	3	0	0
14 iteration_0	plus	PASSED	3	0	0
15 iteration_0	plus	PASSED	3	0	0
16 iteration_0	plus	PASSED	3	0	0
17 iteration_0	plus	PASSED	4	0	0
18 iteration_0	plus	PASSED	3	0	0
19 iteration_0	plus	PASSED	3	0	0
0 iteration_0	test	PASSED	1	0	0
1 iteration_0	test	PASSED	1	0	0
2 iteration_0	test	PASSED	1	0	0
3 iteration_0	test	PASSED	1	0	0
4 iteration_0	test	FAILED	0	1	0
5 iteration_0	test	FAILED	1	0	0
6 iteration_0	test	PASSED	1	0	0
7 iteration_0	test	PASSED	1	0	0
8 iteration_0	test	PASSED	1	0	0
9 iteration_0	test	PASSED	1	0	0
10 iteration_0	test	PASSED	1	0	0
11 iteration_0	test	FAILED	1	0	0
12 iteration_0	test	PASSED	1	0	0
13 iteration_0	test	PASSED	1	0	0
14 iteration_0	test	PASSED	1	0	0
15 iteration_0	test	FAILED	0	1	0
16 iteration_0	test	PASSED	1	0	0
17 iteration_0	test	PASSED	1	0	0
18 iteration_0	test	PASSED	1	0	0
19 iteration_0	test	PASSED	1	0	0
0 iteration_0	test	PASSED	1	0	0
1 iteration_0	test	PASSED	1	0	0
2 iteration_0	test	FAILED	0	1	0
3 iteration_0	test	PASSED	1	0	0
4 iteration_0	test	FAILED	0	1	0
5 iteration_0	test	FAILED	1	0	0
6 iteration_0	test	PASSED	1	0	0

0dcb1eec-3577-46d6-9e64-56a4ad020f7b

7 iteration_0	test	PASSED	1	0	0
8 iteration_0	test	PASSED	1	0	0
9 iteration_0	test	PASSED	1	0	0
10 iteration_0	test	PASSED	1	0	0
11 iteration_0	test	FAILED	1	0	0
12 iteration_0	test	PASSED	1	0	0
13 iteration_0	test	PASSED	1	0	0
14 iteration_0	test	PASSED	1	0	0
15 iteration_0	test	FAILED	1	0	0
16 iteration_0	test	PASSED	1	0	0
17 iteration_0	test	PASSED	1	0	0
18 iteration_0	test	PASSED	1	0	0
19 iteration_0	test	PASSED	1	0	0
20 iteration_0	test	FAILED	0	1	0
21 iteration_0	test	PASSED	1	0	0
22 iteration_0	test	PASSED	1	0	0

passed_private	failed_private	timeout_private	passed_generate	failed_generate	timeout_generate
0	0	0	117	83	0
2	5	0	13	187	0
1	2	0	4	196	0
14	55	0	121	79	0
1	1	0	12	188	0
1	1	0	4	196	0
0	0	1	0	4	1
0	0	0	25	175	0
5	0	0	91	0	0
3	0	0	200	0	0
1	0	0	200	0	0
2	3	0	7	193	0
0	6	1	5	61	1
4	0	0	200	0	0
1	3	0	1	199	0
0	2	0	0	200	0
0	0	0	200	0	0
22	69	0	16	184	0
966	33	0	0	0	0
7	993	0	0	0	0
0	0	0	0	0	0
717	181	0	0	0	0
785	0	0	0	0	0
6	147	0	0	0	0
11	55	0	0	0	0
178	0	0	0	0	0
1000	0	0	0	0	0
265	694	0	0	0	0
50	730	0	0	0	0
405	0	0	0	0	0
580	353	0	0	0	0
987	0	0	0	0	0
753	150	0	0	0	0
134	0	0	0	0	0
286	714	0	0	0	0
215	0	0	0	0	0
1000	0	0	0	0	0
128	0	0	0	0	0
7	0	0	0	0	0
0	0	0	0	0	0
3	0	0	0	0	0
4	2	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

0dcb1eec-3577-46d6-9e64-56a4ad020f7b

3	0	0	0	0	0
3	0	0	0	0	0
5	0	0	0	0	0
5	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
0	0	0	0	0	0
103	0	0	0	0	0
106	0	0	0	0	0
90	18	0	0	0	0
66	36	0	0	0	0
103	0	0	0	0	0
101	0	0	0	0	0
103	0	0	0	0	0
106	0	0	0	0	0
108	0	0	0	0	0
113	0	0	0	0	0
101	0	0	0	0	0
104	0	0	0	0	0
102	0	0	0	0	0
101	0	0	0	0	0
125	0	0	0	0	0
102	0	0	0	0	0
105	0	0	0	0	0
101	0	0	0	0	0
102	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
0	3	0	0	0	0
1	2	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
2	1	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
1	2	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
0	3	0	0	0	0
3	0	0	0	0	0
0	3	0	0	0	0
1	2	0	0	0	0
3	0	0	0	0	0

0dcb1eec-3577-46d6-9e64-56a4ad020f7b

3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
2	1	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
2	1	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0
0	3	0	0	0	0
3	0	0	0	0	0
3	0	0	0	0	0

total_passed	total_failed	total_timeout	total_tests	dataset_name
117	84	0	201	codecontest
17	192	0	209	codecontest
5	201	0	206	codecontest
136	138	0	274	codecontest
13	191	0	204	codecontest
5	199	0	204	codecontest
0	6	2	8	codecontest
26	177	0	203	codecontest
100	0	0	100	codecontest
204	0	0	204	codecontest
202	0	0	202	codecontest
9	198	0	207	codecontest
5	68	2	75	codecontest
205	0	0	205	codecontest
2	203	0	205	codecontest
0	203	0	203	codecontest
201	0	0	201	codecontest
38	254	0	292	codecontest
968	33	0	1001	humaneval
7	994	0	1001	humaneval
0	0	0	0	humaneval
719	181	0	900	humaneval
786	0	0	786	humaneval
7	148	0	155	humaneval
12	55	0	67	humaneval
180	0	0	180	humaneval
1002	0	0	1002	humaneval
266	694	0	960	humaneval
51	732	0	783	humaneval
406	0	0	406	humaneval
583	353	0	936	humaneval
989	0	0	989	humaneval
754	150	0	904	humaneval
136	0	0	136	humaneval
288	714	0	1002	humaneval
216	0	0	216	humaneval
1003	0	0	1003	humaneval
129	0	0	129	humaneval
9	0	0	9	humaneval
0	0	0	0	humaneval
4	0	0	4	humaneval
6	2	0	8	humaneval
4	0	0	4	humaneval
5	0	0	5	humaneval
4	0	0	4	humaneval
6	0	0	6	humaneval
7	0	0	7	humaneval
5	0	0	5	humaneval
8	0	0	8	humaneval
4	0	0	4	humaneval
6	0	0	6	humaneval
6	0	0	6	humaneval

4	0	0	4 humaneval
5	0	0	5 humaneval
7	0	0	7 humaneval
6	0	0	6 humaneval
7	0	0	7 humaneval
6	0	0	6 humaneval
0	0	0	0 mbpp
108	0	0	108 mbpp
109	0	0	109 mbpp
96	18	0	114 mbpp
69	36	0	105 mbpp
106	0	0	106 mbpp
104	0	0	104 mbpp
106	0	0	106 mbpp
109	0	0	109 mbpp
111	0	0	111 mbpp
116	0	0	116 mbpp
104	0	0	104 mbpp
107	0	0	107 mbpp
105	0	0	105 mbpp
104	0	0	104 mbpp
128	0	0	128 mbpp
105	0	0	105 mbpp
109	0	0	109 mbpp
104	0	0	104 mbpp
105	0	0	105 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
0	4	0	4 mbpp
2	2	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
3	1	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
1	3	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
0	4	0	4 mbpp
4	0	0	4 mbpp
0	4	0	4 mbpp
2	2	0	4 mbpp
4	0	0	4 mbpp

0dcb1eec-3577-46d6-9e64-56a4ad020f7b

4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
3	1	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
3	1	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp
0	4	0	4 mbpp
4	0	0	4 mbpp
4	0	0	4 mbpp

results directory

[illegible]

[illegible]

C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results
C:\Users\tayla\OneDrive\Dokumente\Paircoder\A-Pair-Coder\Examples\PairCoder-main\logs\mbpp\test\gpt-4\results