

A4F BENCHMARK ANALYSIS REPORT BASED ON THE RESPONSES OF CHATGPT.

Exp. Settings	Total Instances	Counter Examples	No Counter examples	Instance Found	Overfit (No Instance found)	Syntax error	Type Error	No result / blank Instance
Bug locations and fix comments (Fix bug – No Assertion)	1938	133	497	401	96	637	666	5 (possible reason – No check / Run command in chatGPT solution.)
Bug locations and without Fix comments (Fix bug – No Assertion)	1938	333	668	497	171	372	560	6 (possible reason – No check / Run command in chatGPT solution.)
No Bug locations and No Fix comments (Fix bug – Pass Assertion)	1938	309	415	303	112	590	624	0
No Bug locations and No Fix comments (Fix bug – No Assertion)	1938	137	184	147	37	1310 (Possible Reason) Incomplete responses Partial solutions wrong operators)	307	0
Bug locations and without Fix comments (Fix bug – No Assertion)	1938	218	511	374	137	706	502	1