	5 Models			15 Models			30 Models		
	#Attempts (#Faulty)	Avg(s)	Std(s)	#Attempts	Avg(s)	std(s)	#Attempts	Avg(s)	Std(s)
GPT-3.5-turbo	1.1	6.448	1.2	2.2	6.4486	1.2	4.0	19.648	3.8
GPT-3.5-turbo-16k	1.9	4.702	0.7	2.6	9.729	2.2	2.8	12.974	1.8
GPT-4	1.0	25.729	3.9	1.2	36.254	7.3	1.6	51.561	6.8
Gemini Pro 1.0	1.6	17.435	0.7	2.1	27.393	3.4	3.0	29.012	7.2
Gemini Pro 1.5	1.1	20.193	0.1	1.2	28.386	3.3	1.5	35.940	2.3
Llama3.1	3.7	17.893	2.7	6.0	66.075	23.4	N/A	N/A	N/A
Codex	1.0 (0.0)	6.027	0.8	1.0 (0.0)	8.944	1.3	1.3 (0.16)	21.042	2.1
Qwen2.5-coder	1.7 (0.16)	17.336	3.8	1.2 (0.08)	26.597	2.6	1.2 (0.00)	32.433	4.6
Codegeex4	1.0 (0.0)	20.683	2.4	1.5 (0.4)	30.788	5.8	2.5 (0.5)	34.110	8.9
Codellama	2.3 (0.83)	28.423	6.6	4.5 (1.83)	32.581	7.5	N/A	N/A	N/A