WizardArena Rebuttal

Table 1: The Chinese ELO rankings results of 19 models on LMSYS ChatBot Arena-CH, WizardArena-Diverse-CH, WizardArena-Hard-CH, and WizardArena-Mix-CH (Diverse & Hard).

Model	LMSYS-ChatBot Arena-ELO-CH (95% CI)	WizardArena Diverse-ELO-CH (95% CI)	WizardArena Hard-ELO-CH (95% CI)	WizardArena Mix-ELO-CH (95% CI)
Command R+	1254 (+9/-8)	1366 (+6/-7)	1344 (+5/-6)	1352 (+6/-5)
Qwen1.5-72B-Chat	1241 (+10/-10)	1345 (+6/-5)	1351 (+7/-7)	1349 (+7/-6)
Qwen1.5-32B-Chat	1235 (+13/-12)	1328 (+7/-7)	1312 (+6/-8)	1321 (+7/-8)
Yi-34B-Chat	1232 (+19/-12)	1334 (+6/-4)	1325 (+5/-7)	1328 (+6/-6)
GPT-3.5-Turbo-0613	1157 (+12/-17)	1313 (+7/-8)	1294 (+9/-6)	1302 (+8/-6)
WizardLM-β-DPO-PPO-I1	-	1299 (+6/-8)	1281 (+7/-6)	1288 (+7/-7)
WizardLM-β-PPO-I1	-	1278 (+7/-7)	1267 (+6/-8)	1272 (+6/-8)
OpenChat-3.5	1123 (+31/-35)	1276 (+7/-6)	1263 (+8/-6)	1268 (+7/-5)
WizardLM-β-DPO-I1	-	1271 (+10/-10)	1259 (+8/-7)	1264 (+8/-9)
WizardLM-70B-v1.0	1113 (+35/-29)	1226 (+6/-8)	1218 (+7/-10)	1221 (+6/-9)
Starling-LM-7B-alpha	1107 (+18/-17)	1229 (+7/-6)	1220 (+8/-6)	1225 (+8/-6)
Vicuna-13B	1098 (+21/-18)	1172 (+6/-9)	1158 (+7/-11)	1164 (+6/-10)
Vicuna-33B	1093 (+17/-13)	1199 (+10/-8)	1167 (+9/-10)	1181 (+9/-9)
WizardLM-β-SFT-I1	-	1117 (+11/-8)	1105 (+8/-12)	1112 (+11/-9)
Llama-2-70b-chat	1057 (+14/-17)	1100 (+10/-7)	1100 (+9/-9)	1100 (+10/-8)
Llama-2-13b-Chat	1056 (+20/-18)	1001 (+9/-10)	989 (+11/-11)	993 (+9/-12)
Zephyr-7b-beta	1018 (+30/-30)	871 (+8/-12)	874 (+11/-9)	872 (+10/-10)
Mistral-7B-Instruct-v0.1	990 (+35/-33)	875 (+10/-8)	859 (+8/-7)	868 (+9/-7)
WizardLM-β-SFT-I0	-	812 (+11/-11)	801 (+10/-14)	808 (+10/-12)

Table 2: The ELO rankings results of 26 models on LMSYS ChatBot Arena EN, MT-Bench, Offline-Diverse, Offline-Hard, Offline-Mix (Diverse & Hard). GPT-4o achieved an ELO score of 1397, ranking highest in WizardArena, followed by GPT-4-1106-Preview with 1378 and Claude 3.5 Sonnet with 1386.

Model	LMSYS-ChatBot Arena-ELO-EN (95% CI)	WizardArena Diverse-ELO (95% CI)	WizardArena Hard-ELO (95% CI)	WizardArena Mix-ELO (95% CI)	MT-bench
GPT-4o	1264 (+3/-3)	1402 (+3/-4)	1394 (+4/-5)	1397 (+5/-4)	9.30
Claude 3.5 Sonnet	1245 (+5/-4)	1390 (+4/-6)	1379 (+6/-5)	1386 (+5/-4)	9.20
GPT-4-1106-Preview	1232 (+3/-3)	1374 (+3/-5)	1379 (+6/-4)	1378 (+5/-3)	9.32
Command R+	1163 (+3/-5)	1350 (+9/-7)	1328 (+8/-5)	1339 (+6/-5)	8.20
GPT-4-0613	1147 (+3/-3)	1343 (+6/-7)	1326 (+5/-4)	1334 (+7/-5)	9.18
Qwen1.5-72B-Chat	1137 (+3/-4)	1331 (+8/-8)	1313 (+7/-6)	1323 (+6/-6)	8.61
Qwen1.5-32B-Chat	1115 (+5/-7)	1293 (+7/-8)	1276 (+5/-8)	1286 (+6/-4)	8.30
Wizard-β-PPO-I_3	-	1282 (+6/-6)	1266 (+7/-4)	1274 (+6/-7)	7.81
WizardLM-β-DPO-PPO-I_1	-	1226 (+8/-7)	1204 (+6/-7)	1219 (+7/-5)	7.40
WizardLM-β-PPO-I_1	-	1211 (+7/-8)	1197 (+8/-5)	1205 (+4/-5)	7.29
WizardLM-β-DPO-I_1	-	1203 (+7/-6)	1191 (+7/-9)	1198 (+3/-4)	7.35
WizardLM-70B-v1.0	1099 (+7/-8)	1185 (+7/-6)	1164 (+5/-6)	1172 (+5/-5)	7.71
Tulu-2-DPO-70B	1093 (+8/-10)	1148 (+6/-8)	1178 (+8/-6)	1159 (+5/-7)	7.89
Llama-2-70B-Chat	1091 (+5/-5)	1100 (+6/-5)	1100 (+6/-6)	1100 (+2/-3)	6.86
Vicuna-33B	1088 (+5/-5)	1112 (+6/-7)	1076 (+8/-6)	1092 (+5/-5)	7.12
Nous-Hermes-2-Mixtral-DPO	1079 (+9/-13)	1107 (+8/-6)	1121 (+7/-6)	1116 (+5/-4)	8.33
WizardLM-β-SFT-I_1	-	1064 (+8/-7)	1061 (+7/-6)	1062 (+5/-3)	6.98
OpenChat-3.5	1066 (+7/-7)	1042 (+5/-7)	1050 (+8/-6)	1046 (+5/-5)	7.80
DeepSeek-LLM-67B-Chat	1066 (+8/-10)	991 (+7/-8)	1009 (+5/-7)	1001 (+7/-7)	8.70
Llama-2-13B-Chat	1060 (+4/-5)	1053 (+6/-6)	1041 (+8/-7)	1045 (+5/-4)	6.65
GPT-3.5-Turbo-0613	1055 (+6/-6)	957 (+6/-8)	1007 (+6/-7)	984 (+6/-5)	8.32
Zephyr-7b-alpha	1041 (+14/-15)	907 (+7/-5)	967 (+5/-6)	943 (+4/-4)	6.88
Vicuna-13B	1031 (+5/-6)	934 (+6/-5)	926 (+8/-6)	929 (+5/-5)	6.57
Qwen-14B-Chat	1019 (+9/-10)	917 (+7/-8)	932 (+8/-6)	923 (+4/-6)	6.96
Mistral-7B-Instruct-v0.1	1011 (+7/-7)	884 (+6/-7)	902 (+9/-6)	895 (+6/-5)	6.84
WizardLM-β-SFT-I_0	-	865 (+6/-5)	884 (+8/-7)	873 (+6/-6)	6.41

Table 3: The win/tie/loss counts of WizardLM-β-PPO-I3 against Command R+, Qwen1.5-72B-Chat, OpenChat-3.5 evaluated by Llama3 70B Chat Judge and Human Judge.

Models	Win	Tie	Loss
Llama3 70B Chat Judge			
WizardLM-β-PPO-I3 vs. Command R+	55	39	106
WizardLM-β-PPO-I3 vs. Qwen1.5-72B-Chat	64	45	91
WizardLM-β-PPO-I3 vs. OpenChat-3.5	141	23	36
Human Judge			
WizardLM-β-PPO-I3 vs. Command R+	49	46	105
WizardLM-β-PPO-I3 vs. Qwen1.5-72B-Chat	60	41	99
WizardLM-β-PPO-I3 vs. OpenChat-3.5	147	21	32

Table 4: Explore alignment strategies for models in SFT and RL stages. We utilize three slices of data for SFT, DPO, and PPO training in first round.

Alignment Strategy	Data Source	Offline-Mix Arena-ELO (95% CI)	MT-bench
SFT	D_1	1063 (+4/-4)	6.98
SFT + SFT	$D_1 \cup D_2$	1124 (+6/-4)	7.15
SFT + DPO	$D_1 \cup D_2$	1198 (+3/-4)	7.35
SFT + PPO	$D_1 \cup D_2$	1205 (+4/-5)	7.29
SFT + SFT + SFT	$D_1 \cup D_2 \cup D_3$	1175 (+8/-6)	7.28
SFT + SFT + DPO	$D_1 \cup D_2 \cup D_3$	1204 (+7/-6)	7.32
SFT + SFT + PPO	$D_1 \cup D_2 \cup D_3$	1207 (+5/-7)	7.33
SFT + DPO + DPO	$D_1 \cup D_2 \cup D_3$	1209 (+5/-5)	7.38
SFT + PPO + DPO	$D_1 \cup D_2 \cup D_3$	1213 (+6/-7)	7.34
SFT + DPO + PPO	$D_1 \cup D_2 \cup D_3$	1219 (+7/-5)	7.40

Table 5: The WizardArena Elo of WizardLM-β-7B-SFT-I_1 on different battle modes.

Battle Mode	WizardArena-Mix
i) Ours v.s. OpenChat-3.5	931 (+7/-5)
i) Ours v.s. Qwen-1.5-72B	1023 (+5/-5)
i) Ours v.s. Command R+	1031 (+6/-4)
ii) {Ours, Qwen-1.5-72B, OpenChat-3.5}, 1v.s.1	1034 (+6/-7)
ii) {Ours, Command R+, OpenChat-3.5}, 1v.s.1	1039 (+5/-8)
ii) {Ours, Qwen-1.5-72B, Command R+}, 1v.s.1	1049 (+5/-5)
iii) {Ours, Qwen-1.5-72B, Command R+, OpenChat-3.5}, 1v.s.1	1063 (+4/-4)

Table 6: The performance of WizardLM-β trained on different rounds on WizardArena-Mix.

Training rounds	WizardArena-Mix
WizardLM-β-I1	1219 (+7/-5)
WizardLM-β-I2	1256 (+4/-6)
WizardLM-β-I3	1274 (+6/-7)
WizardLM-β-I4	1281 (+5/-6)

Table 7: The win/tie/loss counts of WizardLM-β-PPO-I3 against GPT-4o, GPT-4-1106-Preview, Claude 3.5 Sonnet evaluated by Llama3 70B Chat Judge and Human Judge. when Llama3-70B-Chat as the judge model, WizardLM-β-I3 achieved win rates of 25.3%, 34.1%, and 28.6% respectively. in human evaluations, WizardLM--PPO-I3 had win rates of 23.8%, 33.3%, and 30.5%.

Models	Win	Tie	Loss
Llama3 70B Chat Judge			
WizardLM-β-PPO-I3 vs. GPT-4o	21	17	62
WizardLM-β-PPO-I3 vs. GPT-4-1106-Preview	29	15	56
WizardLM-β-PPO-I3 vs. Claude 3.5 Sonnet	24	16	60
Human Judge			
WizardLM-β-PPO-I3 vs. GPT-4o	19	20	61
WizardLM-β-PPO-I3 vs. GPT-4-1106-Preview	26	22	52
WizardLM-β-PPO-I3 vs. Claude 3.5 Sonnet	25	18	57

Table 8: The performance impact of employing more advanced models to battle with WizardLM-β-7B-I_0 on different stages.

Training Stage	WizardArena Elo	MT-Bench
SFT-I_0	875 (+5/-5)	6.41
Battles With M_0={Command R+, Qwen1.5-72B-Chat, and OpenChat 3.5}		
SFT-I_1	1063 (+4/-4)	6.98
{SFT + DPO}-I_1	1198 (+3/-4)	7.35
{SFT + DPO + PPO}-I_1	1219 (+7/-5)	7.40
Battles With M_1={GPT-4o, GPT4-1106-Preview, and Claude 3.5 Sonnet}		
SFT-I_1	1164 (+5/-7)	7.21
{SFT + DPO}-I_1	1229 (+7/-6)	7.46
{SFT + DPO + PPO}-I_1	1265 (+6/-5)	7.62

Table 9: The consistency between Llama3-70B-Chat and GPT-4 as judging models in the WizardArena-Mix with 16 models. Using multiple bootstraps (e.g., 100), we select the median as the model's ELO score.

Model	LMSYS-ChatBot Arena-ELO-EN (95% CI)	WizardArena-Mix-ELO GPT-4-judge (95% CI)	WizardArena-Mix-ELO Llama3-70B-chat-judge (95% CI)
Command R+	1163 (+3/-5)	1357 (+4/-3)	1340 (+6/-4)
Qwen1.5-72B-Chat	1137 (+3/-4)	1336 (+5/-5)	1324 (+6/-5)
Qwen1.5-32B-Chat	1115 (+5/-7)	1303 (+4/-6)	1288 (+6/-4)
WizardLM-70B-v1.0	1099 (+7/-8)	1196(+7/-5)	1172 (+5/-5)
Tulu-2-DPO-70B	1093 (+8/-10)	1184(+6/-8)	1161 (+4/-6) -
Llama-2-70B-Chat	1091 (+5/-5)	1100(+4/-5)	1100 (+2/-3) -
Vicuna-33B	1088 (+5/-5)	1089(+7/-6)	1094 (+4/-5)
Nous-Hermes-2-Mixtral-DPO	1079 (+9/-13)	1086(+5/-7)	1116 (+5/-4)
OpenChat-3.5	1066 (+7/-7)	982(+8/-6)	1048 (+5/-5)
DeepSeek-LLM-67B-Chat	1066 (+8/-10)	990 (+6/-8)	1001 (+6/-5)
Llama-2-13B-Chat	1060 (+4/-5)	986(+7/-7)	1047 (+5/-4)
GPT-3.5-Turbo-0613	1055 (+6/-6)	944 (+7/-5)	984 (+5/-5)
Zephyr-7b-alpha	1041 (+14/-15)	932 (+6/-5)	943 (+4/-4)
Vicuna-13B	1031 (+5/-6)	944 (+6/-5)	929 (+5/-5)
Qwen-14B-Chat	1019 (+9/-10)	924 (+4/-7)	926 (+4/-6)
Mistral-7B-Instruct-v0.1	1011 (+7/-7)	887(+6/-9)	895 (+4/-5)