Table 1: Bias and Accuracy Results for Non-Control Categories (short answer) (Language = EN)

Category	Llama-2-7b			Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis	Acc	BiasAmb	BiasDis
Age	0.304	0.835	0.376	0.610	0.720	0.432
Disability_status	0.295	0.865	0.343	0.616	0.655	0.459
Gender_identity	0.322	0.821	0.294	0.605	0.641	0.372
Physical_appearance	0.303	0.872	0.377	0.541	0.697	0.368
SES	0.340	0.849	0.357	0.692	0.633	0.436
Sexual_orientation	0.331	0.839	0.378	0.605	0.694	0.467
ALL	0.317	0.847	0.361	0.632	0.673	0.426

Table 2: Bias and Accuracy Results for Control Categories (short answer) (Language = EN)

Category	Llama-2-7b			Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis	Acc	BiasAmb	BiasDis
Age_control	0.311	0.818	0.388	0.688	0.636	0.428
Disability_status_control	0.328	0.787	0.429	0.716	0.624	0.459
Gender_identity_control	0.323	0.865	0.414	0.663	0.660	0.349
Physical_appearance_control	0.307	0.820	0.422	0.698	0.621	0.388
SES_control	0.350	0.843	0.392	0.774	0.604	0.431
Sexual_orientation_control	0.320	0.796	0.408	0.658	0.630	0.306
ALL	0.327	0.826	0.400	0.722	0.622	0.422

Table 3: Bias and Accuracy Results for Non-Control Categories (cot) (Language = EN)

Category	Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis
Age	0.333	0.818	0.240
Disability_status	0.341	0.824	0.224
Gender_identity	0.341	0.832	0.222
Physical_appearance	0.313	0.786	0.191
SES	0.334	0.789	0.274
Sexual_orientation	0.296	0.839	0.267
ALL	0.332	0.806	0.243

Table 4: Bias and Accuracy Results for Control Categories (cot) (Language = EN)

Category	Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis
Age_control	0.345	0.800	0.199
Disability_status_control	0.299	0.818	0.200
Gender_identity_control	0.354	0.744	0.220
Physical_appearance_control	0.327	0.789	0.207
SES_control	0.346	0.776	0.245
Sexual_orientation_control	0.276	0.778	0.204
ALL	0.337	0.789	0.218

Table 5: Bias and Accuracy Results for Non-Control Categories (reasoning) (Language = EN)

Category	DeepSeek-R1-8B		
	Acc	BiasAmb	BiasDis
Age	0.552	0.723	0.371
Disability_status	0.565	0.735	0.392
Gender_identity	0.516	0.674	0.372
Physical_appearance	0.550	0.714	0.360
SES	0.565	0.695	0.414
Sexual_orientation	0.520	0.694	0.311
ALL	0.556	0.710	0.387

Table 6: Bias and Accuracy Results for Control Categories (reasoning) (Language = EN)

Category	DeepSeek-R1-8B		
	Acc	BiasAmb	BiasDis
Age_control			
Disability_status_control	0.610	0.668	0.385
Gender_identity_control	0.617	0.635	0.392
Physical_appearance_control	0.583	0.708	0.356
SES_control	0.613	0.665	0.403
Sexual_orientation_control	0.658	0.741	0.388
ALL	0.608	0.672	0.390

Table 7: Bias and Accuracy Results for Non-Control Categories (short answer) (Language = ES)

Category	Llama-2-7b			Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis	Acc	BiasAmb	BiasDis
Age	0.283	0.862	0.377	0.537	0.729	0.419
Disability_status	0.300	0.831	0.319	0.533	0.687	0.422
Gender_identity	0.292	0.837	0.394	0.500	0.690	0.367
Physical_appearance	0.304	0.903	0.385	0.531	0.675	0.392
SES	0.322	0.878	0.372	0.685	0.649	0.472
Sexual_orientation	0.257	0.887	0.244	0.559	0.710	0.400
ALL	0.302	0.867	0.368	0.587	0.686	0.432

Table 8: Bias and Accuracy Results for Control Categories (short answer) (Language = ES)

Category	Llama-2-7b			Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis	Acc	BiasAmb	BiasDis
Age_control	0.306	0.851	0.390	0.605	0.661	0.420
Disability_status_control	0.323	0.851	0.400	0.591	0.681	0.418
Gender_identity_control	0.330	0.887	0.391	0.595	0.675	0.418
Physical_appearance_control	0.295	0.902	0.380	0.638	0.638	0.403
SES_control	0.343	0.842	0.418	0.734	0.608	0.464
Sexual_orientation_control	0.342	0.768	0.417	0.625	0.625	0.438
ALL	0.322	0.854	0.400	0.653	0.642	0.434

Table 9: Bias and Accuracy Results for Non-Control Categories (cot) (Language = ES)

Category	Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis
Age	0.333	0.730	0.282
Disability_status	0.342	0.712	0.287
Gender_identity	0.328	0.712	0.306
Physical_appearance	0.331	0.689	0.279
SES	0.325	0.722	0.320
Sexual_orientation	0.368	0.742	0.156
ALL	0.331	0.720	0.295

Table 10: Bias and Accuracy Results for Control Categories (cot) (Language = ES)

Category	Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis
Age_control	0.335	0.734	0.273
Disability_status_control	0.326	0.693	0.273
Gender_identity_control	0.348	0.725	0.272
Physical_appearance_control	0.340	0.726	0.285
SES_control	0.332	0.709	0.291
Sexual_orientation_control	0.316	0.679	0.292
ALL	0.334	0.717	0.281

Table 11: Bias and Accuracy Results for Non-Control Categories (reasoning) (Language = ES)

Category	DeepSeek-R1-8B		
	Acc	BiasAmb	BiasDis
Age	0.512	0.634	0.414
Disability_status	0.495	0.600	0.422
Gender_identity	0.471	0.712	0.450
Physical_appearance	0.523	0.642	0.434
SES	0.586	0.598	0.457
Sexual_orientation	0.500	0.645	0.356
ALL	0.535	0.622	0.434

Table 12: Bias and Accuracy Results for Control Categories (reasoning) (Language = ES)

Category	DeepSeek-R1-8B		
	Acc	BiasAmb	BiasDis
Age_control	0.516	0.672	0.420
Disability_status_control	0.544	0.637	0.462
Gender_identity_control	0.574	0.625	0.429
Physical_appearance_control	0.589	0.601	0.438
SES_control	0.607	0.624	0.456
Sexual_orientation_control	0.572	0.589	0.500
ALL	0.565	0.638	0.442

Table 13: Bias and Accuracy Results for Non-Control Categories (short answer) (Language = TR)

Category	Llama-2-7b			Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis	Acc	BiasAmb	BiasDis
Age	0.295	0.976	0.028	0.416	0.759	0.347
Disability_status	0.289	0.986	0.047	0.373	0.739	0.365
Gender_identity	0.300	0.984	0.056	0.395	0.763	0.335
Physical_appearance	0.292	0.972	0.037	0.409	0.804	0.388
SES	0.293	0.982	0.071	0.509	0.698	0.410
Sexual_orientation	0.257	0.952	0.000	0.487	0.613	0.311
ALL	0.293	0.979	0.048	0.443	0.737	0.376

Table 14: Bias and Accuracy Results for Control Categories (short answer) (Language = TR)

Category	Llama-2-7b			Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis	Acc	BiasAmb	BiasDis
Age_control	0.310	0.974	0.045	0.438	0.703	0.397
Disability_status_control	0.287	0.977	0.062	0.431	0.696	0.376
Gender_identity_control	0.309	0.987	0.032	0.396	0.684	0.389
Physical_appearance_control	0.290	0.983	0.056	0.448	0.690	0.406
SES_control	0.301	0.975	0.087	0.532	0.683	0.447
Sexual_orientation_control	0.316	1.000	0.021	0.447	0.741	0.298
ALL	0.301	0.977	0.062	0.470	0.693	0.411

Table 15: Bias and Accuracy Results for Non-Control Categories (cot) (Language = TR)

Category	Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis
Age	0.330	0.728	0.322
Disability_status	0.338	0.677	0.326
Gender_identity	0.355	0.726	0.341
Physical_appearance	0.323	0.688	0.292
SES	0.351	0.702	0.308
Sexual_orientation	0.408	0.677	0.222
ALL	0.340	0.706	0.314

Table 16: Bias and Accuracy Results for Control Categories (cot) (Language = TR)

Category	Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis
Age_control	0.342	0.691	0.308
Disability_status_control	0.346	0.738	0.295
Gender_identity_control	0.332	0.759	0.297
Physical_appearance_control	0.333	0.659	0.311
SES_control	0.335	0.695	0.298
Sexual_orientation_control	0.329	0.638	0.298
ALL	0.338	0.697	0.302

Table 17: Bias and Accuracy Results for Non-Control Categories (reasoning) (Language = TR)

Category	DeepSeek-R1-8B		
	Acc	BiasAmb	BiasDis
Age	0.382	0.670	0.356
Disability_status	0.355	0.729	0.358
Gender_identity	0.375	0.694	0.430
Physical_appearance	0.346	0.749	0.366
SES	0.422	0.713	0.403
Sexual_orientation	0.368	0.710	0.311
ALL	0.388	0.704	0.378

Table 18: Bias and Accuracy Results for Control Categories (reasoning) (Language = TR)

Category	DeepSeek-R1-8B		
	Acc	BiasAmb	BiasDis
Age_control	0.380	0.664	0.388
Disability_status_control	0.400	0.692	0.376
Gender_identity_control	0.347	0.690	0.449
Physical_appearance_control	0.384	0.662	0.406
SES_control	0.418	0.692	0.417
Sexual_orientation_control	0.388	0.603	0.447
ALL	0.395	0.678	0.403

Table 19: Bias and Accuracy Results for Non-Control Categories (short answer) (Language = NL)

Category	Llama-2-7b			Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis	Acc	BiasAmb	BiasDis
Age	0.284	0.878	0.356	0.544	0.641	0.473
Disability_status	0.291	0.858	0.375	0.533	0.653	0.445
Gender_identity	0.302	0.853	0.389	0.524	0.641	0.444
Physical_appearance	0.301	0.900	0.397	0.507	0.681	0.424
SES	0.313	0.876	0.397	0.704	0.576	0.491
Sexual_orientation	0.270	0.855	0.289	0.618	0.597	0.489
ALL	0.298	0.876	0.379	0.595	0.622	0.469

Table 20: Bias and Accuracy Results for Control Categories (short answer) (Language = NL)

Category	Llama-2-7b			Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis	Acc	BiasAmb	BiasDis
Age_control	0.285	0.887	0.367	0.571	0.591	0.459
Disability_status_control	0.275	0.907	0.410	0.632	0.579	0.459
Gender_identity_control	0.298	0.882	0.441	0.542	0.625	0.399
Physical_appearance_control	0.295	0.901	0.442	0.641	0.614	0.422
SES_control	0.318	0.892	0.394	0.719	0.592	0.477
Sexual_orientation_control	0.296	0.942	0.540	0.566	0.615	0.360
ALL	0.297	0.894	0.398	0.638	0.594	0.456

Table 21: Bias and Accuracy Results for Non-Control Categories (cot) (Language = NL)

Category	Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis
Age	0.339	0.711	0.277
Disability_status	0.336	0.715	0.245
Gender_identity	0.335	0.663	0.306
Physical_appearance	0.337	0.739	0.289
SES	0.333	0.703	0.285
Sexual_orientation	0.329	0.710	0.333
ALL	0.336	0.709	0.279

Table 22: Bias and Accuracy Results for Control Categories (cot) (Language = NL)

Category	Llama-3.1-8B-Instruct		
	Acc	BiasAmb	BiasDis
Age_control	0.339	0.722	0.288
Disability_status_control	0.348	0.713	0.339
Gender_identity_control	0.316	0.697	0.266
Physical_appearance_control	0.328	0.730	0.272
SES_control	0.337	0.697	0.285
Sexual_orientation_control	0.303	0.673	0.280
ALL	0.336	0.710	0.290

Table 23: Bias and Accuracy Results for Non-Control Categories (reasoning) (Language = NL)

Category	DeepSeek-R1-8B		
	Acc	BiasAmb	BiasDis
Age	0.433	0.679	0.451
Disability_status	0.417	0.680	0.448
Gender_identity	0.399	0.625	0.333
Physical_appearance	0.442	0.639	0.409
SES	0.482	0.650	0.465
Sexual_orientation	0.434	0.597	0.400
ALL	0.448	0.660	0.443

Table 24: Bias and Accuracy Results for Control Categories (reasoning) (Language = NL)

Category	DeepSeek-R1-8B		
	Acc	BiasAmb	BiasDis
Age_control	0.423	0.676	0.424
Disability_status_control	0.428	0.694	0.486
Gender_identity_control	0.430	0.658	0.468
Physical_appearance_control	0.462	0.639	0.425
SES_control	0.458	0.662	0.450
Sexual_orientation_control	0.487	0.596	0.420
ALL	0.442	0.667	0.444