

Bias in Question Answering on BBQ by Identity Characteristic: Disambiguated Contexts

Source: Parrish et al., 2022; Glaese et al., 2022 | Chart: 2023 AI Index Report

Category	Age	-3.00	2.70	4.40	2.40	3.30	1.20	7.00	8.00
	Disability Status	5.40	5.70	8.10	1.70	-0.70	-1.40	0.00	8.00
	Gender Identity	14.00	2.90	4.60	-16.90	-3.40	-5.80	2.00	3.00
	Gender Identity (Names)	-0.90	1.10	3.60	0.40	2.00	0.10		
	Nationality	-0.10	0.70	5.70	1.90	-0.20	1.20	-2.00	3.00
	Physical Appearance	17.10	-2.70	4.20	-5.00	-1.70	-2.30	12.00	8.00
	Race/Ethnicity	0.60	-0.80	1.20	0.00	0.90	0.00	3.00	1.00
	Race/Ethnicity (Names)	0.40	-0.20	-0.30	0.00	0.30	-0.10		
	Religion	5.20	3.40	1.80	1.70	3.50	0.20	5.00	7.00
	Sexual Orientation	6.50	-3.10	-4.80	-0.20	0.50	-0.70	-1.00	-1.00
	Socio-Economic Status	7.00	3.50	3.80	2.90	3.80	3.90	8.00	7.00
			RoBERTa-Base	RoBERTa-Large	DeBERTaV3-Base	DeBERTaV3-Large	UnifiedQA (ARC)	UnifiedQA (RACE)	Dialogue-Prompted Chinchilla (DPC)
Model									

