

Supplementary Material

In the following tables, the proportion of the generated texts with a specific output attribute (e.g., conservative) for a given demographic group (e.g., male) is reported with one, two, or three asterisks when binomial testing indicates that the proportion is significantly different from the real-world, ground-truth statistic at the 0.05, 0.01, or 0.001 level, respectively. In each table, the Cohen's h effect size is also reported in parentheses, and the 95% Wilson confidence interval is reported below each proportion^{49,50}. When the real-world, ground-truth statistic for a proportion is not available, the Cohen's h effect size is reported as "N/A". The proportion of the generated texts with a specific output attribute (e.g., conservative) for a given demographic group (e.g., male) is also reported with one, two, or three daggers when Fisher's exact test indicates that the proportion for the texts differs significantly between implicit and explicit prompt types at the 0.05, 0.01, or 0.001 level, respectively[?].

Politics Tables

Implicit

		claude-3.5-sonnet			
		Conservative	Liberal	Neutral	Refusal
Gender	Male (n=500)	4.20*** ($h=-0.70$) [2.76, 6.34]	93.80*** ($h=1.57$) [91.33, 95.60]	2.00*** ($h=-1.13$) [1.09, 3.64]	0.00
	Female (n=500)	9.20*** ($h=-0.43$) [7.00, 12.10] [†]	90.00*** ($h=1.16$) [87.45, 92.65] [†]	0.40*** ($h=-1.08$) [0.11, 1.45]	0.40
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-1.07$) [0.00, 7.41]	92.00*** ($h=1.51$) [86.02, 98.85]	4.00*** ($h=-0.90$) [1.15, 13.98]	4.00
	White (n=50)	18.00* ($h=-0.32$) [10.42, 32.54]	74.00*** ($h=1.11$) [65.10, 88.01]	2.00*** ($h=-1.01$) [0.38, 11.11]	6.00
	Black (n=50)	12.00** ($h=0.37$) [5.86, 24.70]	80.00*** ($h=0.95$) [70.42, 91.30] [†]	4.00*** ($h=-0.68$) [1.15, 13.98]	4.00
	Hispanic (n=50)	0.00*** ($h=-0.77$) [0.00, 7.13]	96.00*** ($h=1.39$) [86.54, 98.90]	4.00*** ($h=-0.91$) [1.10, 13.46]	0.00
	Asian (n=50)	0.00** ($h=-0.71$) [0.00, 7.87]	90.00*** ($h=1.69$) [92.13, 100.00]	0.00*** ($h=-1.41$) [0.00, 7.87]	10.00
Age	Baby Boomer (n=50)	26.00* ($h=-0.30$) [15.87, 39.55] ^{†††}	70.00*** ($h=0.94$) [56.25, 80.90] ^{†††}	4.00*** ($h=-0.82$) [1.10, 13.46]	0.00
	Generation X (n=50)	0.00*** ($h=-1.29$) [0.00, 7.13]	100.00*** ($h=2.09$) [92.87, 100.00] ^{††}	0.00*** ($h=-1.33$) [0.00, 7.13] [†]	0.00
	Millennial (n=50)	2.00*** ($h=-0.74$) [0.35, 10.50]	98.00*** ($h=1.51$) [89.50, 99.65]	0.00*** ($h=-1.29$) [0.00, 7.13]	0.00
	Generation Z (n=50)	0.00*** ($h=-1.12$) [0.00, 7.13]	100.00*** ($h=1.71$) [92.87, 100.00]	0.00*** ($h=-1.12$) [0.00, 7.13]	0.00
	Generation Alpha (n=50)	8.00 ($h=N/A$) [3.15, 18.84]	92.00 ($h=N/A$) [81.16, 96.85]	0.00 ($h=N/A$) [0.00, 7.13]	0.00

Table A1. Politics analysis of implicit bias for claude-3.5-sonnet.

gpt-4o-mini					
		Conservative	Liberal	Neutral	Refusal
Gender	Male (n=500)	0.20*** ($h=-1.03$) [0.04, 1.12]	99.80*** ($h=1.98$) [98.88, 99.96]	0.00*** ($h=-1.41$) [0.00, 0.76]	0.00
	Female (n=500)	0.00*** ($h=-1.05$) [0.00, 0.76]	100.00*** ($h=1.79$) [99.24, 100.00]	0.00*** ($h=-1.20$) [0.00, 0.76]	0.00
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-1.07$) [0.00, 7.13]	100.00*** ($h=1.92$) [92.87, 100.00]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	White (n=50)	0.00*** ($h=-1.22$) [0.00, 7.13]	100.00*** ($h=2.07$) [92.87, 100.00]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	Black (n=50)	0.00 ($h=-0.35$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.09$) [0.00, 7.13]	0.00
	Hispanic (n=50)	2.00** ($h=-0.48$) [0.35, 10.50]	98.00*** ($h=1.51$) [89.50, 99.65]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	Asian (n=50)	0.00** ($h=-0.71$) [0.00, 7.13]	100.00*** ($h=1.69$) [92.87, 100.00]	0.00*** ($h=-1.41$) [0.00, 7.13]	0.00
Age	Baby Boomer (n=50)	2.00*** ($h=-1.09$) [0.35, 10.50] [†]	98.00*** ($h=1.81$) [89.50, 99.65] [†]	0.00*** ($h=-1.22$) [0.00, 7.13]	0.00
	Generation X (n=50)	0.00*** ($h=-1.29$) [0.00, 7.13]	100.00*** ($h=2.09$) [92.87, 100.00]	0.00*** ($h=-1.33$) [0.00, 7.13]	0.00
	Millennial (n=50)	0.00*** ($h=-1.02$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.29$) [0.00, 7.13]	0.00
	Generation Z (n=50)	0.00*** ($h=-1.12$) [0.00, 7.13]	100.00*** ($h=1.71$) [92.87, 100.00]	0.00*** ($h=-1.12$) [0.00, 7.13]	0.00
	Generation Alpha (n=50)	0.00 ($h=N/A$) [0.00, 7.13]	100.00 ($h=N/A$) [92.87, 100.00]	0.00 ($h=N/A$) [0.00, 7.13]	0.00

Table A2. Politics analysis of implicit bias for gpt-4o-mini.

llama-3.1-70b					
		Conservative	Liberal	Neutral	Refusal
Gender	Male (n=500)	1.40*** ($h=-0.88$) [0.68, 2.86]	98.60*** ($h=1.83$) [97.14, 99.32]	0.00*** ($h=-1.41$) [0.00, 0.76]	0.00
	Female (n=500)	0.80*** ($h=-0.87$) [0.31, 2.04]	99.20*** ($h=1.61$) [97.96, 99.69]	0.00*** ($h=-1.20$) [0.00, 0.76]	0.00
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-1.07$) [0.00, 7.13]	100.00*** ($h=1.92$) [92.87, 100.00]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	White (n=50)	8.00*** ($h=-0.65$) [3.15, 18.84]	92.00*** ($h=1.50$) [81.16, 96.85]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	Black (n=50)	2.00 ($h=-0.06$) [0.35, 10.50]	98.00*** ($h=1.51$) [89.50, 99.65]	0.00*** ($h=-1.09$) [0.00, 7.13]	0.00
	Hispanic (n=50)	0.00*** ($h=-0.77$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	Asian (n=50)	0.00** ($h=-0.71$) [0.00, 7.13]	88.00*** ($h=0.98$) [76.20, 94.38] [†]	12.00*** ($h=-0.70$) [5.62, 23.80] [†]	0.00
	Baby Boomer (n=50)	2.00*** ($h=-1.09$) [0.35, 10.50] ^{†††}	98.00*** ($h=1.81$) [89.50, 99.65] ^{†††}	0.00*** ($h=-1.22$) [0.00, 7.13]	0.00
Age	Generation X (n=50)	0.00*** ($h=-1.29$) [0.00, 7.13]	100.00*** ($h=2.09$) [92.87, 100.00]	0.00*** ($h=-1.33$) [0.00, 7.13]	0.00
	Millennial (n=50)	0.00*** ($h=-1.02$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.29$) [0.00, 7.13]	0.00
	Generation Z (n=50)	0.00*** ($h=-1.12$) [0.00, 7.13]	100.00*** ($h=1.71$) [92.87, 100.00]	0.00*** ($h=-1.12$) [0.00, 7.13]	0.00
	Generation Alpha (n=50)	8.00 ($h=N/A$) [3.15, 18.84]	92.00 ($h=N/A$) [81.16, 96.85]	0.00 ($h=N/A$) [0.00, 7.13]	0.00

Table A3. Politics analysis of implicit bias for llama-3.1-70b.

command-r-plus					
		Conservative	Liberal	Neutral	Refusal
Gender	Male (n=500)	24.60 ($h=-0.07$) [21.22, 28.77] ^{†††}	70.60*** ($h=0.93$) [66.46, 74.42] ^{†††}	4.80*** ($h=-0.98$) [3.08, 6.81]	0.00
	Female (n=500)	4.40*** ($h=-0.62$) [2.92, 6.57] ^{†††}	95.20*** ($h=1.35$) [92.96, 96.75] ^{†††}	0.40*** ($h=-1.08$) [0.11, 1.45]	0.00
Ethnicity/Race	Neutral (n=50)	12.00* ($h=-0.36$) [5.62, 23.80]	88.00*** ($h=1.21$) [76.20, 94.38] [†]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	White (n=50)	14.00** ($h=-0.46$) [6.95, 26.19] ^{†††}	78.00*** ($h=1.10$) [64.76, 87.25] ^{†††}	8.00*** ($h=-0.73$) [3.15, 18.84]	0.00
	Black (n=50)	0.00 ($h=-0.35$) [0.00, 7.13]	94.00*** ($h=1.30$) [83.78, 97.94]	6.00*** ($h=-0.60$) [2.06, 16.22]	0.00
	Hispanic (n=50)	8.00 ($h=-0.19$) [3.15, 18.84]	92.00*** ($h=1.22$) [81.16, 96.85]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	Asian (n=50)	12.00*** ($h=0.00$) [5.62, 23.80]	82.00*** ($h=0.81$) [69.20, 90.23]	6.00*** ($h=-0.92$) [2.06, 16.22]	0.00
Age	Baby Boomer (n=50)	12.00*** ($h=-0.66$) [5.62, 23.80] ^{†††}	86.00*** ($h=1.33$) [73.81, 93.05] ^{†††}	2.00*** ($h=-0.94$) [0.35, 10.50]	0.00
	Generation X (n=50)	4.00*** ($h=-0.88$) [1.10, 13.46] ^{†††}	92.00*** ($h=1.52$) [81.16, 96.85] ^{†††}	4.00*** ($h=-0.93$) [1.10, 13.46] [†]	0.00
	Millennial (n=50)	10.00* ($h=-0.38$) [4.35, 21.36]	88.00*** ($h=1.09$) [76.20, 94.38]	2.00*** ($h=-1.00$) [0.35, 10.50]	0.00
	Generation Z (n=50)	6.00*** ($h=-0.62$) [2.06, 16.22]	90.00*** ($h=1.07$) [78.64, 95.65]	4.00*** ($h=-0.71$) [1.10, 13.46]	0.00
	Generation Alpha (n=50)	6.00 ($h=N/A$) [2.06, 16.22]	94.00 ($h=N/A$) [83.78, 97.94]	0.00 ($h=N/A$) [0.00, 7.13]	0.00

Table A4. Politics analysis of implicit bias for command-r-plus.

Explicit

		claude-3.5-sonnet			
		Conservative	Liberal	Neutral	Refusal
Gender	Male (n=50)	0.00*** ($h=-1.12$) [0.00, 7.13]	100.00*** ($h=2.07$) [92.87, 100.00]	0.00*** ($h=-1.41$) [0.00, 7.13]	0.00
	Female (n=50)	0.00*** ($h=-1.05$) [0.00, 7.13] [†]	100.00*** ($h=1.79$) [92.87, 100.00] [†]	0.00*** ($h=-1.20$) [0.00, 7.13]	0.00
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-1.07$) [0.00, 7.13]	100.00*** ($h=1.92$) [92.87, 100.00]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	White (n=50)	0.00*** ($h=N/A$) [0.00, 0.00]	0.00*** ($h=N/A$) [0.00, 0.00]	0.00*** ($h=N/A$) [0.00, 0.00]	100.00
	Black (n=50)	0.00 ($h=-0.35$) [0.00, 13.32]	50.00*** ($h=1.79$) [86.68, 100.00] [†]	0.00*** ($h=-1.09$) [0.00, 13.32]	50.00
	Hispanic (n=50)	0.00** ($h=-0.77$) [0.00, 7.41]	96.00*** ($h=1.79$) [92.59, 100.00]	0.00*** ($h=-1.31$) [0.00, 7.41]	4.00
	Asian (n=50)	0.00 ($h=-0.71$) [0.00, 12.87]	48.00*** ($h=1.13$) [75.86, 97.86]	4.00*** ($h=-0.85$) [2.14, 24.14]	48.00
Age	Baby Boomer (n=50)	90.00*** ($h=1.13$) [78.64, 95.65] ^{†††}	8.00** ($h=-0.47$) [3.15, 18.84] ^{†††}	2.00*** ($h=-0.94$) [0.35, 10.50]	0.00
	Generation X (n=50)	6.00*** ($h=-0.79$) [2.06, 16.22]	82.00*** ($h=1.22$) [69.20, 90.23] ^{††}	12.00*** ($h=-0.62$) [5.62, 23.80] [†]	0.00
	Millennial (n=50)	0.00*** ($h=-1.02$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.29$) [0.00, 7.13]	0.00
	Generation Z (n=50)	0.00*** ($h=-1.12$) [0.00, 7.13]	100.00*** ($h=1.71$) [92.87, 100.00]	0.00*** ($h=-1.12$) [0.00, 7.13]	0.00
	Generation Alpha (n=50)	0.00 ($h=N/A$) [0.00, 7.13]	100.00 ($h=N/A$) [92.87, 100.00]	0.00 ($h=N/A$) [0.00, 7.13]	0.00

Table A5. Politics analysis of explicit bias for claude-3.5-sonnet.

gpt-4o-mini					
		Conservative	Liberal	Neutral	Refusal
Gender	Male (n=50)	0.00*** ($h=-1.12$) [0.00, 7.13]	100.00*** ($h=2.07$) [92.87, 100.00]	0.00*** ($h=-1.41$) [0.00, 7.13]	0.00
	Female (n=50)	0.00*** ($h=-1.05$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.20$) [0.00, 7.13]	0.00
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-1.07$) [0.00, 7.13]	100.00*** ($h=1.92$) [92.87, 100.00]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	White (n=50)	4.00*** ($h=-0.82$) [1.10, 13.46]	96.00*** ($h=1.67$) [86.54, 98.90]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	Black (n=50)	0.00 ($h=-0.35$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.09$) [0.00, 7.13]	0.00
	Hispanic (n=50)	4.00* ($h=-0.36$) [1.10, 13.46]	96.00*** ($h=1.39$) [86.54, 98.90]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	Asian (n=50)	0.00** ($h=-0.71$) [0.00, 7.13]	100.00*** ($h=1.69$) [92.87, 100.00]	0.00*** ($h=-1.41$) [0.00, 7.13]	0.00
Age	Baby Boomer (n=50)	18.00** ($h=-0.49$) [9.77, 30.80] [†]	82.00*** ($h=1.22$) [69.20, 90.23] [†]	0.00*** ($h=-1.22$) [0.00, 7.13]	0.00
	Generation X (n=50)	2.00*** ($h=-1.00$) [0.35, 10.50]	98.00*** ($h=1.81$) [89.50, 99.65]	0.00*** ($h=-1.33$) [0.00, 7.13]	0.00
	Millennial (n=50)	0.00*** ($h=-1.02$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.29$) [0.00, 7.13]	0.00
	Generation Z (n=50)	0.00*** ($h=-1.12$) [0.00, 7.13]	100.00*** ($h=1.71$) [92.87, 100.00]	0.00*** ($h=-1.12$) [0.00, 7.13]	0.00
	Generation Alpha (n=50)	0.00 ($h=N/A$) [0.00, 7.13]	100.00 ($h=N/A$) [92.87, 100.00]	0.00 ($h=N/A$) [0.00, 7.13]	0.00

Table A6. Politics analysis of explicit bias for gpt-4o-mini.

llama-3.1-70b					
		Conservative	Liberal	Neutral	Refusal
Gender	Male (n=50)	2.00*** ($h=-0.83$) [0.35, 10.50]	98.00*** ($h=1.79$) [89.50, 99.65]	0.00*** ($h=-1.41$) [0.00, 7.13]	0.00
	Female (n=50)	0.00*** ($h=-1.05$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.20$) [0.00, 7.13]	0.00
Ethnicity/Race	Neutral (n=50)	2.00*** ($h=-0.79$) [0.35, 10.50]	98.00*** ($h=1.63$) [89.50, 99.65]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	White (n=50)	8.00*** ($h=-0.65$) [3.15, 18.84]	92.00*** ($h=1.50$) [81.16, 96.85]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	Black (n=50)	0.00 ($h=-0.35$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.09$) [0.00, 7.13]	0.00
	Hispanic (n=50)	0.00*** ($h=-0.77$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.31$) [0.00, 7.13]	0.00
	Asian (n=50)	0.00** ($h=-0.71$) [0.00, 7.13]	100.00*** ($h=1.69$) [92.87, 100.00] [†]	0.00*** ($h=-1.41$) [0.00, 7.13] [†]	0.00
Age	Baby Boomer (n=50)	100.00*** ($h=1.77$) [92.87, 100.00] ^{†††}	0.00*** ($h=-1.05$) [0.00, 7.13] ^{†††}	0.00*** ($h=-1.22$) [0.00, 7.13]	0.00
	Generation X (n=50)	4.00*** ($h=-0.88$) [1.10, 13.46]	94.00*** ($h=1.60$) [83.78, 97.94]	2.00*** ($h=-1.04$) [0.35, 10.50]	0.00
	Millennial (n=50)	0.00*** ($h=-1.02$) [0.00, 7.13]	100.00*** ($h=1.79$) [92.87, 100.00]	0.00*** ($h=-1.29$) [0.00, 7.13]	0.00
	Generation Z (n=50)	0.00*** ($h=-1.12$) [0.00, 7.13]	100.00*** ($h=1.71$) [92.87, 100.00]	0.00*** ($h=-1.12$) [0.00, 7.13]	0.00
	Generation Alpha (n=50)	0.00 ($h=N/A$) [0.00, 7.13]	96.00 ($h=N/A$) [86.54, 98.90]	4.00 ($h=N/A$) [1.10, 13.46]	0.00

Table A7. Politics analysis of explicit bias for llama-3.1-70b.

command-r-plus					
		Conservative	Liberal	Neutral	Refusal
Gender	Male (n=50)	70.00*** ($h=0.87$) [56.25, 80.90] ^{†††}	24.00 ($h=-0.05$) [14.30, 37.41] ^{†††}	6.00*** ($h=-0.92$) [2.06, 16.22]	0.00
	Female (n=50)	18.00 ($h=-0.17$) [9.77, 30.80] ^{†††}	80.00*** ($h=0.87$) [66.96, 88.76] ^{†††}	2.00*** ($h=-0.92$) [0.35, 10.50]	0.00
Ethnicity/Race	Neutral (n=50)	20.00 ($h=-0.14$) [11.24, 33.04]	70.00*** ($h=0.76$) [56.25, 80.90] [†]	10.00*** ($h=-0.66$) [4.35, 21.36]	0.00
	White (n=50)	60.00*** ($h=0.55$) [46.18, 72.39] ^{†††}	38.00 ($h=0.26$) [25.86, 51.85] ^{†††}	2.00*** ($h=-1.02$) [0.35, 10.50]	0.00
	Black (n=50)	0.00 ($h=-0.35$) [0.00, 7.13]	98.00*** ($h=1.51$) [89.50, 99.65]	2.00*** ($h=-0.81$) [0.35, 10.50]	0.00
	Hispanic (n=50)	16.00 ($h=0.06$) [8.34, 28.51]	80.00*** ($h=0.87$) [66.96, 88.76]	4.00*** ($h=-0.91$) [1.10, 13.46]	0.00
	Asian (n=50)	2.00* ($h=-0.42$) [0.35, 10.50]	90.00*** ($h=1.05$) [78.64, 95.65]	8.00*** ($h=-0.84$) [3.15, 18.84]	0.00
Age	Baby Boomer (n=50)	78.00*** ($h=0.80$) [64.76, 87.25] ^{†††}	16.00 ($h=-0.22$) [8.34, 28.51] ^{†††}	6.00*** ($h=-0.73$) [2.06, 16.22]	0.00
	Generation X (n=50)	26.00 ($h=-0.22$) [15.87, 39.55] ^{††}	54.00*** ($h=0.60$) [40.40, 67.03] ^{†††}	20.00** ($h=-0.40$) [11.24, 33.04] [†]	0.00
	Millennial (n=50)	0.00*** ($h=-1.02$) [0.00, 7.13]	96.00*** ($h=1.39$) [86.54, 98.90]	4.00*** ($h=-0.88$) [1.10, 13.46]	0.00
	Generation Z (n=50)	0.00*** ($h=-1.12$) [0.00, 7.13]	96.00*** ($h=1.31$) [86.54, 98.90]	4.00*** ($h=-0.71$) [1.10, 13.46]	0.00
	Generation Alpha (n=50)	2.00 ($h=N/A$) [0.35, 10.50]	98.00 ($h=N/A$) [89.50, 99.65]	0.00 ($h=N/A$) [0.00, 7.13]	0.00

Table A8. Politics analysis of explicit bias for command-r-plus.

Religion Tables
Implicit

		claude-3.5-sonnet						
		Buddhist	Christian	Hindu	Jewish	Muslim	Unaffiliated	Refusal
Gender	Male (n=500)	0.60 ($h=-0.05$) [0.20, 1.75]	6.20*** ($h=-1.41$) [4.40, 8.67]	0.00* ($h=-0.20$) [0.00, 0.76]	1.80 ($h=-0.01$) [0.95, 3.39]	0.00* ($h=-0.20$) [0.00, 0.76]	91.40*** ($h=1.45$) [88.62, 93.55]	0.00
	Female (n=500)	1.00 ($h=0.00$) [0.43, 2.33] ^{†††}	43.00*** ($h=-0.68$) [38.89, 47.56] ^{†††}	0.60 ($h=-0.04$) [0.21, 1.76]	2.00 ($h=0.00$) [1.09, 3.66]	0.00* ($h=-0.20$) [0.00, 0.77]	53.00*** ($h=0.73$) [48.82, 57.55] ^{†††}	0.40
Ethnicity/Race	Neutral (n=50)	0.00 ($h=-0.20$) [0.00, 7.41] [†]	12.00*** ($h=-1.26$) [5.86, 24.70]	2.00 ($h=0.09$) [0.37, 10.90]	10.00** ($h=0.37$) [4.53, 22.17] [†]	0.00 ($h=-0.20$) [0.00, 7.41]	72.00*** ($h=1.05$) [61.22, 85.08]	4.00
	White (n=50)	0.00 ($h=-0.20$) [-0.00, 7.56]	0.00*** ($h=-2.05$) [-0.00, 7.56]	10.00*** ($h=0.46$) [4.63, 22.59]	82.00*** ($h=2.06$) [74.83, 94.02]	2.00 ($h=0.09$) [0.38, 11.11]	0.00*** ($h=-1.02$) [-0.00, 7.56]	6.00
	Black (n=50)	6.00* ($h=0.31$) [2.15, 16.84]	10.00*** ($h=-1.58$) [4.53, 22.17] ^{†††}	4.00 ($h=0.21$) [1.15, 13.98]	0.00 ($h=-0.20$) [0.00, 7.41]	50.00*** ($h=1.33$) [38.33, 65.53] ^{†††}	26.00 ($h=0.22$) [16.57, 41.00]	4.00
	Hispanic (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	22.00*** ($h=-1.19$) [12.75, 35.24] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	4.00 ($h=0.20$) [1.10, 13.46]	74.00*** ($h=1.14$) [60.45, 84.13] ^{††}	0.00
	Asian (n=50)	28.00*** ($h=0.98$) [19.53, 45.66]	16.00*** ($h=-0.46$) [9.29, 31.33] [†]	4.00* ($h=-0.40$) [1.23, 14.83]	0.00 ($h=-0.49$) [0.00, 7.87]	18.00** ($h=0.43$) [10.90, 33.82] [†]	24.00 ($h=-0.10$) [15.96, 41.04] ^{†††}	10.00
Age	Baby Boomer (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	50.00*** ($h=-0.62$) [36.64, 63.36] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	8.00* ($h=0.29$) [3.15, 18.84]	0.00 ($h=-0.20$) [0.00, 7.13]	42.00*** ($h=0.56$) [29.38, 55.77] ^{†††}	0.00
	Generation X (n=50)	4.00 ($h=0.20$) [1.10, 13.46]	24.00*** ($h=-0.98$) [14.30, 37.41] ^{†††}	2.00 ($h=0.08$) [0.35, 10.50]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	70.00*** ($h=0.98$) [56.25, 80.90] ^{†††}	0.00
	Millennial (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	8.00*** ($h=-1.16$) [3.15, 18.84]	0.00 ($h=-0.25$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.25$) [0.00, 7.13]	92.00*** ($h=1.30$) [81.16, 96.85]	0.00
	Generation Z (n=50)	2.00 ($h=0.08$) [0.35, 10.50]	12.00*** ($h=-0.98$) [5.62, 23.80] [†]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	2.00 ($h=0.00$) [0.35, 10.50]	84.00*** ($h=1.07$) [71.49, 91.66] ^{††}	0.00
	Generation Alpha (n=50)	2.00 ($h=0.08$) [0.35, 10.50]	6.00*** ($h=-1.40$) [2.06, 16.22]	8.00** ($h=0.37$) [3.15, 18.84]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.35$) [0.00, 7.13]	84.00*** ($h=1.29$) [71.49, 91.66] ^{††}	0.00

Table A9. Religion analysis of implicit bias for claude-3.5-sonnet.

gpt-4o-mini								
		Buddhist	Christian	Hindu	Jewish	Muslim	Unaffiliated	Refusal
Gender	Male (n=500)	2.00* ($h=0.08$) [1.09, 3.64]	53.60*** ($h=-0.27$) [49.22, 57.93]	0.00* ($h=-0.20$) [0.00, 0.76]	0.00*** ($h=-0.28$) [0.00, 0.76]	1.00 ($h=0.00$) [0.43, 2.32]	43.40*** ($h=0.35$) [39.12, 47.78]	0.00
	Female (n=500)	0.60 ($h=-0.05$) [0.20, 1.75] ^{††}	86.40*** ($h=0.27$) [83.12, 89.13] ^{††}	0.00* ($h=-0.20$) [0.00, 0.76]	0.20** ($h=-0.19$) [0.04, 1.12]	0.00* ($h=-0.20$) [0.00, 0.76]	12.80*** ($h=-0.17$) [10.15, 16.01]	0.00
Ethnicity/Race	Neutral (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	62.00 ($h=-0.17$) [48.15, 74.14]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	38.00* ($h=0.28$) [25.86, 51.85] [†]	0.00
	White (n=50)	2.00 ($h=0.08$) [0.35, 10.50]	4.00*** ($h=-1.65$) [1.10, 13.46] ^{†††}	2.00 ($h=0.08$) [0.35, 10.50]	80.00*** ($h=1.87$) [66.96, 88.76] ^{†††}	8.00** ($h=0.37$) [3.15, 18.84]	4.00*** ($h=-0.62$) [1.10, 13.46]	0.00
	Black (n=50)	2.00 ($h=0.08$) [0.35, 10.50]	30.00*** ($h=-1.08$) [19.10, 43.75] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	46.00*** ($h=1.21$) [32.97, 59.60] ^{†††}	22.00 ($h=0.10$) [12.75, 35.24] ^{†††}	0.00
	Hispanic (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	74.00 ($h=-0.09$) [60.45, 84.13] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	10.00*** ($h=0.44$) [4.35, 21.36]	16.00 ($h=-0.10$) [8.34, 28.51] ^{††}	0.00
	Asian (n=50)	20.00*** ($h=0.73$) [11.24, 33.04] ^{†††}	36.00 ($h=-0.04$) [24.14, 49.86] ^{†††}	0.00*** ($h=-0.82$) [0.00, 7.13]	0.00 ($h=-0.49$) [0.00, 7.13]	30.00*** ($h=0.66$) [19.10, 43.75] ^{†††}	14.00** ($h=-0.41$) [6.95, 26.19]	0.00
Age	Baby Boomer (n=50)	2.00 ($h=0.08$) [0.35, 10.50]	86.00 ($h=0.19$) [73.81, 93.05] [†]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	12.00 ($h=-0.14$) [5.62, 23.80] [†]	0.00
	Generation X (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	58.00 ($h=-0.27$) [44.23, 70.62]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	42.00** ($h=0.41$) [29.38, 55.77]	0.00
	Millennial (n=50)	2.00 ($h=0.08$) [0.35, 10.50]	60.00 ($h=0.04$) [46.18, 72.39] ^{†††}	0.00 ($h=-0.25$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.25$) [0.00, 7.13]	38.00 ($h=0.06$) [25.86, 51.85] ^{†††}	0.00
	Generation Z (n=50)	2.00 ($h=0.08$) [0.35, 10.50]	54.00 ($h=-0.04$) [40.40, 67.03] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	44.00 ($h=0.21$) [31.16, 57.69] ^{†††}	0.00
	Generation Alpha (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	46.00** ($h=-0.41$) [32.97, 59.60] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.35$) [0.00, 7.13]	54.00*** ($h=0.63$) [40.40, 67.03] ^{†††}	0.00

Table A10. Religion analysis of implicit bias for gpt-4o-mini.

llama-3.1-70b								
		Buddhist	Christian	Hindu	Jewish	Muslim	Unaffiliated	Refusal
Gender	Male (n=500)	19.40*** ($h=0.71$) [16.17, 23.09]	52.60*** ($h=-0.29$) [48.22, 56.94]	0.00* ($h=-0.20$) [0.00, 0.76]	0.40** ($h=-0.16$) [0.11, 1.45]	0.00* ($h=-0.20$) [0.00, 0.76]	27.60 ($h=0.01$) [23.86, 31.68]	0.00
	Female (n=500)	26.60*** ($h=0.88$) [22.92, 30.64] ^{†††}	54.20*** ($h=-0.46$) [49.82, 58.52] ^{†††}	0.00* ($h=-0.20$) [0.00, 0.76]	0.20** ($h=-0.19$) [0.04, 1.12]	0.00* ($h=-0.20$) [0.00, 0.76]	19.00 ($h=0.00$) [15.80, 22.67]	0.00
Ethnicity/Race	Neutral (n=50)	38.00*** ($h=1.13$) [25.86, 51.85]	26.00*** ($h=-0.91$) [15.87, 39.55]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	36.00 ($h=0.24$) [24.14, 49.86] [†]	0.00
	White (n=50)	14.00*** ($h=0.57$) [6.95, 26.19] [†]	0.00*** ($h=-2.05$) [0.00, 7.13] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	78.00*** ($h=1.82$) [64.76, 87.25] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	8.00** ($h=-0.45$) [3.15, 18.84]	0.00
	Black (n=50)	34.00*** ($h=1.04$) [22.44, 47.85] ^{†††}	10.00*** ($h=-1.60$) [4.35, 21.36] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	40.00*** ($h=1.09$) [27.61, 53.82] ^{†††}	16.00 ($h=-0.05$) [8.34, 28.51] ^{††}	0.00
	Hispanic (n=50)	22.00*** ($h=0.78$) [12.75, 35.24] ^{†††}	44.00*** ($h=-0.71$) [31.16, 57.69] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	2.00 ($h=0.08$) [0.35, 10.50]	32.00* ($h=0.28$) [20.76, 45.81] ^{†††}	0.00
	Asian (n=50)	46.00*** ($h=1.29$) [32.97, 59.60] ^{†††}	0.00*** ($h=-1.33$) [0.00, 7.13]	0.00*** ($h=-0.82$) [0.00, 7.13]	0.00 ($h=-0.49$) [0.00, 7.13]	30.00*** ($h=0.66$) [19.10, 43.75] ^{†††}	24.00 ($h=-0.16$) [14.30, 37.41] ^{†††}	0.00
	Baby Boomer (n=50)	22.00*** ($h=0.78$) [12.75, 35.24] ^{†††}	66.00* ($h=-0.29$) [52.15, 77.56] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	12.00 ($h=-0.14$) [5.62, 23.80] [†]	0.00
	Generation X (n=50)	40.00*** ($h=1.17$) [27.61, 53.82] ^{†††}	16.00*** ($h=-1.18$) [8.34, 28.51] [†]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	44.00** ($h=0.45$) [31.16, 57.69]	0.00
Age	Millennial (n=50)	32.00*** ($h=1.00$) [20.76, 45.81]	18.00*** ($h=-0.86$) [9.77, 30.80]	0.00 ($h=-0.25$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.25$) [0.00, 7.13]	50.00* ($h=0.30$) [36.64, 63.36]	0.00
	Generation Z (n=50)	40.00*** ($h=1.17$) [27.61, 53.82]	6.00*** ($h=-1.20$) [2.06, 16.22]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	54.00** ($h=0.41$) [40.40, 67.03]	0.00
	Generation Alpha (n=50)	50.00*** ($h=1.37$) [36.64, 63.36] ^{††}	16.00*** ($h=-1.07$) [8.34, 28.51] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.35$) [0.00, 7.13]	34.00 ($h=0.22$) [22.44, 47.85] ^{††}	0.00

Table A11. Religion analysis of implicit bias for llama-3.1-70b.

command-r-plus								
		Buddhist	Christian	Hindu	Jewish	Muslim	Unaffiliated	Refusal
Gender	Male (n=500)	0.20 ($h=-0.11$) [0.04, 1.12]	69.20 ($h=0.05$) [65.02, 73.09] ^{††}	0.00* ($h=-0.20$) [0.00, 0.76]	0.00*** ($h=-0.28$) [0.00, 0.76]	0.40 ($h=-0.07$) [0.11, 1.45]	30.20 ($h=0.07$) [26.34, 34.36] ^{††}	0.00
	Female (n=500)	1.20 ($h=0.02$) [0.55, 2.59]	69.20*** ($h=-0.15$) [65.02, 73.09]	0.20 ($h=-0.11$) [0.04, 1.12]	0.20** ($h=-0.19$) [0.04, 1.12]	0.20 ($h=-0.11$) [0.04, 1.12]	29.00*** ($h=0.24$) [25.19, 33.13]	0.00
Ethnicity/Race	Neutral (n=50)	4.00 ($h=0.20$) [1.10, 13.46]	60.00 ($h=-0.21$) [46.18, 72.39]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	36.00 ($h=0.24$) [24.14, 49.86]	0.00
	White (n=50)	4.00 ($h=0.20$) [1.10, 13.46]	2.00*** ($h=-1.76$) [0.35, 10.50] ^{†††}	8.00** ($h=0.37$) [3.15, 18.84]	72.00*** ($h=1.68$) [58.33, 82.53] ^{†††}	2.00 ($h=0.08$) [0.35, 10.50]	12.00* ($h=-0.32$) [5.62, 23.80]	0.00
	Black (n=50)	4.00 ($h=0.20$) [1.10, 13.46]	20.00*** ($h=-1.31$) [11.24, 33.04] ^{†††}	4.00 ($h=0.20$) [1.10, 13.46]	0.00 ($h=-0.20$) [0.00, 7.13]	44.00*** ($h=1.17$) [31.16, 57.69] ^{†††}	28.00 ($h=0.24$) [17.47, 41.67] ^{††}	0.00
	Hispanic (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	72.00 ($h=-0.14$) [58.33, 82.53]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	2.00 ($h=0.08$) [0.35, 10.50]	26.00 ($h=0.14$) [15.87, 39.55]	0.00
	Asian (n=50)	20.00*** ($h=0.73$) [11.24, 33.04] ^{†††}	34.00 ($h=-0.08$) [22.44, 47.85] ^{††}	2.00** ($h=-0.54$) [0.35, 10.50]	0.00 ($h=-0.49$) [0.00, 7.13]	28.00*** ($h=0.62$) [17.47, 41.67] ^{†††}	16.00* ($h=-0.36$) [8.34, 28.51]	0.00
	Baby Boomer (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	80.00 ($h=0.02$) [66.96, 88.76] ^{††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	20.00 ($h=0.08$) [11.24, 33.04] ^{††}	0.00
	Generation X (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	56.00* ($h=-0.31$) [42.31, 68.84]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	44.00** ($h=0.45$) [31.16, 57.69]	0.00
Age	Millennial (n=50)	4.00 ($h=0.20$) [1.10, 13.46]	60.00 ($h=0.04$) [46.18, 72.39] [†]	0.00 ($h=-0.25$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.25$) [0.00, 7.13]	36.00 ($h=0.02$) [24.14, 49.86] ^{††}	0.00
	Generation Z (n=50)	6.00* ($h=0.29$) [2.06, 16.22]	50.00 ($h=-0.12$) [36.64, 63.36] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	44.00 ($h=0.21$) [31.16, 57.69] ^{†††}	0.00
	Generation Alpha (n=50)	4.00 ($h=0.20$) [1.10, 13.46]	58.00 ($h=-0.17$) [44.23, 70.62] ^{††}	2.00 ($h=0.08$) [0.35, 10.50]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.35$) [0.00, 7.13]	36.00 ($h=0.26$) [24.14, 49.86] ^{†††}	0.00

Table A12. Religion analysis of implicit bias for command-r-plus.

Explicit

		claude-3.5-sonnet						
		Buddhist	Christian	Hindu	Jewish	Muslim	Unaffiliated	Refusal
Gender	Male (n=50)	4.00 ($h=0.20$) [1.10, 13.46]	0.00*** ($h=-1.92$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	96.00*** ($h=1.65$) [86.54, 98.90]	0.00
	Female (n=50)	10.00*** ($h=0.44$) [4.35, 21.36] ^{†††}	2.00*** ($h=-1.83$) [0.35, 10.50] ^{†††}	2.00 ($h=0.08$) [0.35, 10.50]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	86.00*** ($h=1.47$) [73.81, 93.05] ^{†††}	0.00
Ethnicity/Race	Neutral (n=50)	14.00*** ($h=0.57$) [6.95, 26.19] [†]	2.00*** ($h=-1.70$) [0.35, 10.50]	2.00 ($h=0.08$) [0.35, 10.50]	0.00 ($h=-0.28$) [0.00, 7.13] [†]	0.00 ($h=-0.20$) [0.00, 7.13]	82.00*** ($h=1.22$) [69.20, 90.23]	0.00
	White (n=50)	0.00*** ($h=N/A$) [0.00, 0.00]	0.00*** ($h=N/A$) [0.00, 0.00]	0.00*** ($h=N/A$) [0.00, 0.00]	0.00*** ($h=N/A$) [0.00, 0.00]	0.00*** ($h=N/A$) [0.00, 0.00]	0.00*** ($h=N/A$) [0.00, 0.00]	100.00
	Black (n=50)	0.00 ($h=-0.20$) [0.00, 13.32]	30.00* ($h=-0.47$) [40.74, 76.60] ^{†††}	0.00 ($h=-0.20$) [0.00, 13.32]	0.00 ($h=-0.20$) [0.00, 13.32]	0.00 ($h=-0.28$) [0.00, 13.32]	20.00** ($h=0.49$) [23.40, 59.26]	50.00
	Hispanic (n=50)	0.00 ($h=-0.20$) [0.00, 7.41]	56.00** ($h=-0.43$) [44.28, 71.15] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.41]	0.00 ($h=-0.20$) [0.00, 7.41]	0.00 ($h=-0.20$) [0.00, 7.41]	40.00*** ($h=0.48$) [28.85, 55.72] ^{††}	4.00
	Asian (n=50)	12.00*** ($h=0.80$) [11.03, 42.05]	0.00*** ($h=-1.33$) [0.00, 12.87] [†]	2.00 ($h=-0.43$) [0.68, 18.89]	0.00 ($h=-0.49$) [0.00, 12.87]	0.00 ($h=-0.49$) [0.00, 12.87] [†]	38.00*** ($h=0.87$) [53.92, 86.30] ^{†††}	48.00
Age	Baby Boomer (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	100.00*** ($h=0.95$) [92.87, 100.00] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-0.85$) [0.00, 7.13] ^{†††}	0.00
	Generation X (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-2.00$) [0.00, 7.13] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	100.00*** ($h=2.14$) [92.87, 100.00] ^{†††}	0.00
	Millennial (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-1.73$) [0.00, 7.13]	0.00 ($h=-0.25$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.25$) [0.00, 7.13]	100.00*** ($h=1.88$) [92.87, 100.00]	0.00
	Generation Z (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-1.69$) [0.00, 7.13] [†]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	100.00*** ($h=1.90$) [92.87, 100.00] ^{††}	0.00
	Generation Alpha (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-1.90$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.35$) [0.00, 7.13]	100.00*** ($h=2.12$) [92.87, 100.00] ^{††}	0.00

Table A13. Religion analysis of explicit bias for claude-3.5-sonnet.

gpt-4o-mini								
		Buddhist	Christian	Hindu	Jewish	Muslim	Unaffiliated	Refusal
Gender	Male (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	60.00 ($h=-0.15$) [46.18, 72.39]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	40.00 ($h=0.28$) [27.61, 53.82]	0.00
	Female (n=50)	8.00** ($h=0.37$) [3.15, 18.84] ^{††}	70.00 ($h=-0.14$) [56.25, 80.90] ^{††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	22.00 ($h=0.07$) [12.75, 35.24]	0.00
Ethnicity/Race	Neutral (n=50)	4.00 ($h=0.20$) [1.10, 13.46]	78.00 ($h=0.18$) [64.76, 87.25]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	18.00 ($h=-0.17$) [9.77, 30.80] [†]	0.00
	White (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	100.00*** ($h=1.09$) [92.87, 100.00] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.35$) [0.00, 7.13] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-1.02$) [0.00, 7.13]	0.00
	Black (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	100.00*** ($h=0.90$) [92.87, 100.00] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13] ^{†††}	0.00*** ($h=-0.88$) [0.00, 7.13] ^{†††}	0.00
	Hispanic (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	100.00*** ($h=0.98$) [92.87, 100.00] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-0.93$) [0.00, 7.13] ^{††}	0.00
	Asian (n=50)	98.00*** ($h=2.66$) [89.50, 99.65] ^{†††}	0.00*** ($h=-1.33$) [0.00, 7.13] ^{†††}	0.00*** ($h=-0.82$) [0.00, 7.13]	0.00 ($h=-0.49$) [0.00, 7.13]	0.00 ($h=-0.49$) [0.00, 7.13] ^{†††}	2.00*** ($h=-0.90$) [0.35, 10.50]	0.00
Age	Baby Boomer (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	100.00*** ($h=0.95$) [92.87, 100.00] [†]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-0.85$) [0.00, 7.13] [†]	0.00
	Generation X (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	52.00** ($h=-0.39$) [38.51, 65.20]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	48.00*** ($h=0.53$) [34.80, 61.49]	0.00
	Millennial (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-1.73$) [0.00, 7.13] ^{†††}	0.00 ($h=-0.25$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.25$) [0.00, 7.13]	100.00*** ($h=1.88$) [92.87, 100.00] ^{†††}	0.00
	Generation Z (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-1.69$) [0.00, 7.13] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	100.00*** ($h=1.90$) [92.87, 100.00] ^{†††}	0.00
	Generation Alpha (n=50)	4.00 ($h=0.20$) [1.10, 13.46]	2.00*** ($h=-1.61$) [0.35, 10.50] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.35$) [0.00, 7.13]	94.00*** ($h=1.62$) [83.78, 97.94] ^{†††}	0.00

Table A14. Religion analysis of explicit bias for gpt-4o-mini.

llama-3.1-70b								
		Buddhist	Christian	Hindu	Jewish	Muslim	Unaffiliated	Refusal
Gender	Male (n=50)	18.00*** ($h=0.68$) [9.77, 30.80]	54.00 ($h=-0.27$) [40.40, 67.03]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	28.00 ($h=0.02$) [17.47, 41.67]	0.00
	Female (n=50)	72.00*** ($h=1.83$) [58.33, 82.53] ^{†††}	18.00*** ($h=-1.24$) [9.77, 30.80] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	10.00 ($h=-0.26$) [4.35, 21.36]	0.00
Ethnicity/Race	Neutral (n=50)	50.00*** ($h=1.37$) [36.64, 63.36]	36.00*** ($h=-0.70$) [24.14, 49.86]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	14.00 ($h=-0.28$) [6.95, 26.19] [†]	0.00
	White (n=50)	0.00 ($h=-0.20$) [0.00, 7.13] [†]	84.00 ($h=0.27$) [71.49, 91.66] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.35$) [0.00, 7.13] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	16.00 ($h=-0.20$) [8.34, 28.51]	0.00
	Black (n=50)	0.00 ($h=-0.20$) [0.00, 7.13] ^{†††}	100.00*** ($h=0.90$) [92.87, 100.00] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13] ^{†††}	0.00*** ($h=-0.88$) [0.00, 7.13] ^{††}	0.00
	Hispanic (n=50)	0.00 ($h=-0.20$) [0.00, 7.13] ^{†††}	100.00*** ($h=0.98$) [92.87, 100.00] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-0.93$) [0.00, 7.13] ^{†††}	0.00
	Asian (n=50)	94.00*** ($h=2.45$) [83.78, 97.94] ^{†††}	0.00*** ($h=-1.33$) [0.00, 7.13]	6.00 ($h=-0.33$) [2.06, 16.22]	0.00 ($h=-0.49$) [0.00, 7.13]	0.00 ($h=-0.49$) [0.00, 7.13] ^{†††}	0.00*** ($h=-1.18$) [0.00, 7.13] ^{†††}	0.00
Age	Baby Boomer (n=50)	0.00 ($h=-0.20$) [0.00, 7.13] ^{†††}	100.00*** ($h=0.95$) [92.87, 100.00] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-0.85$) [0.00, 7.13] [†]	0.00
	Generation X (n=50)	4.00 ($h=0.20$) [1.10, 13.46] ^{†††}	38.00*** ($h=-0.68$) [25.86, 51.85] [†]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	58.00*** ($h=0.73$) [44.23, 70.62]	0.00
	Millennial (n=50)	28.00*** ($h=0.91$) [17.47, 41.67]	8.00*** ($h=-1.16$) [3.15, 18.84]	0.00 ($h=-0.25$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.25$) [0.00, 7.13]	64.00*** ($h=0.59$) [50.14, 75.86]	0.00
	Generation Z (n=50)	54.00*** ($h=1.45$) [40.40, 67.03]	4.00*** ($h=-1.29$) [1.10, 13.46]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	42.00 ($h=0.17$) [29.38, 55.77]	0.00
	Generation Alpha (n=50)	18.00*** ($h=0.68$) [9.77, 30.80] ^{††}	70.00 ($h=0.09$) [56.25, 80.90] ^{†††}	2.00 ($h=0.08$) [0.35, 10.50]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.35$) [0.00, 7.13]	10.00* ($h=-0.38$) [4.35, 21.36] ^{††}	0.00

Table A15. Religion analysis of explicit bias for llama-3.1-70b.

command-r-plus								
		Buddhist	Christian	Hindu	Jewish	Muslim	Unaffiliated	Refusal
Gender	Male (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	88.00** ($h=0.52$) [76.20, 94.38] ^{††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	12.00* ($h=-0.39$) [5.62, 23.80] ^{††}	0.00
	Female (n=50)	4.00 ($h=0.20$) [1.10, 13.46]	62.00* ($h=-0.30$) [48.15, 74.14]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	34.00* ($h=0.34$) [22.44, 47.85]	0.00
Ethnicity/Race	Neutral (n=50)	4.00 ($h=0.20$) [1.10, 13.46]	76.00 ($h=0.14$) [62.59, 85.70]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	2.00 ($h=0.08$) [0.35, 10.50]	18.00 ($h=-0.17$) [9.77, 30.80]	0.00
	White (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	80.00 ($h=0.17$) [66.96, 88.76] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.35$) [0.00, 7.13] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	20.00 ($h=-0.10$) [11.24, 33.04]	0.00
	Black (n=50)	2.00 ($h=0.08$) [0.35, 10.50]	92.00* ($h=0.33$) [81.16, 96.85] ^{†††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	2.00 ($h=0.00$) [0.35, 10.50] ^{†††}	4.00** ($h=-0.47$) [1.10, 13.46] ^{††}	0.00
	Hispanic (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	84.00 ($h=0.15$) [71.49, 91.66]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	16.00 ($h=-0.10$) [8.34, 28.51]	0.00
	Asian (n=50)	62.00*** ($h=1.61$) [48.15, 74.14] ^{†††}	10.00*** ($h=-0.68$) [4.35, 21.36] ^{††}	0.00*** ($h=-0.82$) [0.00, 7.13]	0.00 ($h=-0.49$) [0.00, 7.13]	0.00 ($h=-0.49$) [0.00, 7.13] ^{†††}	28.00 ($h=-0.07$) [17.47, 41.67]	0.00
Age	Baby Boomer (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	100.00*** ($h=0.95$) [92.87, 100.00] ^{††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00*** ($h=-0.85$) [0.00, 7.13] ^{††}	0.00
	Generation X (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	54.00* ($h=-0.35$) [40.40, 67.03]	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.20$) [0.00, 7.13]	46.00*** ($h=0.49$) [32.97, 59.60]	0.00
	Millennial (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	34.00*** ($h=-0.49$) [22.44, 47.85] [†]	0.00 ($h=-0.25$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.25$) [0.00, 7.13]	66.00*** ($h=0.63$) [52.15, 77.56] ^{††}	0.00
	Generation Z (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	16.00*** ($h=-0.87$) [8.34, 28.51] ^{†††}	2.00 ($h=0.08$) [0.35, 10.50]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	82.00*** ($h=1.02$) [69.20, 90.23] ^{†††}	0.00
	Generation Alpha (n=50)	0.00 ($h=-0.20$) [0.00, 7.13]	26.00*** ($h=-0.83$) [15.87, 39.55] ^{††}	0.00 ($h=-0.20$) [0.00, 7.13]	0.00 ($h=-0.28$) [0.00, 7.13]	0.00 ($h=-0.35$) [0.00, 7.13]	74.00*** ($h=1.05$) [60.45, 84.13] ^{†††}	0.00

Table A16. Religion analysis of explicit bias for command-r-plus.

Sexual Orientation Tables
Implicit

		claude-3.5-sonnet					
		Heterosexual	LGBTQ	Homosexual	Bisexual	Other	Refusal
Gender	Male (n=500)	5.80*** ($h=-2.31$) [4.07, 8.21]	94.20*** ($h=2.31$) [91.79, 95.93]	88.80	5.40	0.00	0.00
	Female (n=500)	29.00*** ($h=-1.65$) [25.30, 33.25] ^{†††}	70.60*** ($h=1.72$) [66.75, 74.70] ^{†††}	1.80	68.80	0.00	0.40
Ethnicity/Race	Neutral (n=50)	4.00*** ($h=-2.17$) [1.15, 13.98]	92.00*** ($h=2.17$) [86.02, 98.85]	48.00	44.00	0.00	4.00
	White (n=50)	16.00*** ($h=-1.79$) [8.89, 30.14]	78.00*** ($h=1.79$) [69.86, 91.11]	32.00	46.00	0.00	6.00
	Black (n=50)	32.00*** ($h=-1.39$) [21.68, 47.46]	64.00*** ($h=1.39$) [52.54, 78.32]	14.00	50.00	0.00	4.00
	Hispanic (n=50)	0.00*** ($h=-2.47$) [0.00, 7.13] ^{†††}	100.00*** ($h=2.47$) [92.87, 100.00] ^{†††}	46.00	54.00	0.00	0.00
	Asian (n=50)	16.00*** ($h=-1.88$) [9.29, 31.33] [†]	74.00*** ($h=1.88$) [68.67, 90.71] [†]	26.00	48.00	0.00	10.00
Age	Baby Boomer (n=50)	46.00*** ($h=-0.94$) [32.97, 59.60] ^{†††}	54.00*** ($h=1.12$) [40.40, 67.03] ^{†††}	38.00	16.00	0.00	0.00
	Generation X (n=50)	6.00*** ($h=-1.80$) [2.06, 16.22] ^{†††}	94.00*** ($h=1.85$) [83.78, 97.94] ^{†††}	26.00	68.00	0.00	0.00
	Millennial (n=50)	4.00*** ($h=-1.65$) [1.10, 13.46]	96.00*** ($h=1.79$) [86.54, 98.90]	40.00	56.00	0.00	0.00
	Generation Z (n=50)	0.00*** ($h=-1.83$) [0.00, 7.13]	100.00*** ($h=1.94$) [92.87, 100.00]	38.00	62.00	0.00	0.00
	Generation Alpha (n=50)	10.00 ($h=N/A$) [4.35, 21.36]	90.00 ($h=N/A$) [78.64, 95.65]	28.00	62.00	0.00	0.00

Table A17. Sexual orientation analysis of implicit bias for claude-3.5-sonnet.

gpt-4o-mini							
		Heterosexual	LGBTQ	Homosexual	Bisexual	Other	Refusal
Gender	Male (n=500)	7.00*** ($h=-2.26$) [5.08, 9.58] [†]	93.00*** ($h=2.26$) [90.42, 94.92] [†]	60.00	33.00	0.00	0.00
	Female (n=500)	10.60*** ($h=-2.13$) [8.20, 13.61] ^{††}	89.40*** ($h=2.19$) [86.39, 91.80] ^{††}	0.00	89.40	0.00	0.00
Ethnicity/Race	Neutral (n=50)	2.00*** ($h=-2.30$) [0.35, 10.50]	98.00*** ($h=2.30$) [89.50, 99.65]	22.00	76.00	0.00	0.00
	White (n=50)	18.00*** ($h=-1.76$) [9.77, 30.80] ^{†††}	82.00*** ($h=1.76$) [69.20, 90.23] ^{†††}	30.00	52.00	0.00	0.00
	Black (n=50)	12.00*** ($h=-1.91$) [5.62, 23.80]	88.00*** ($h=1.91$) [76.20, 94.38]	12.00	76.00	0.00	0.00
	Hispanic (n=50)	8.00*** ($h=-1.89$) [3.15, 18.84] [†]	92.00*** ($h=1.89$) [81.16, 96.85] [†]	30.00	62.00	0.00	0.00
	Asian (n=50)	14.00*** ($h=-1.98$) [6.95, 26.19]	86.00*** ($h=1.98$) [73.81, 93.05]	16.00	70.00	0.00	0.00
Age	Baby Boomer (n=50)	18.00*** ($h=-1.56$) [9.77, 30.80] ^{†††}	82.00*** ($h=1.73$) [69.20, 90.23] ^{†††}	34.00	48.00	0.00	0.00
	Generation X (n=50)	6.00*** ($h=-1.80$) [2.06, 16.22] ^{†††}	94.00*** ($h=1.85$) [83.78, 97.94] ^{†††}	22.00	72.00	0.00	0.00
	Millennial (n=50)	10.00*** ($h=-1.41$) [4.35, 21.36]	90.00*** ($h=1.55$) [78.64, 95.65]	30.00	60.00	0.00	0.00
	Generation Z (n=50)	0.00*** ($h=-1.83$) [0.00, 7.13]	100.00*** ($h=1.94$) [92.87, 100.00]	20.00	80.00	0.00	0.00
	Generation Alpha (n=50)	12.00 ($h=N/A$) [5.62, 23.80]	88.00 ($h=N/A$) [76.20, 94.38]	4.00	84.00	0.00	0.00

Table A18. Sexual orientation analysis of implicit bias for gpt-4o-mini.

llama-3.1-70b							
		Heterosexual	LGBTQ	Homosexual	Bisexual	Other	Refusal
Gender	Male (n=500)	1.20*** ($h=-2.57$) [0.55, 2.59]	98.80*** ($h=2.57$) [97.41, 99.45]	79.60	19.20	0.00	0.00
	Female (n=500)	4.20*** ($h=-2.38$) [2.76, 6.34]	95.80*** ($h=2.44$) [93.66, 97.24]	14.80	80.80	0.20	0.00
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-2.58$) [0.00, 7.13]	100.00*** ($h=2.58$) [92.87, 100.00]	44.00	56.00	0.00	0.00
	White (n=50)	10.00*** ($h=-1.99$) [4.35, 21.36] ^{††}	90.00*** ($h=1.99$) [78.64, 95.65] ^{††}	46.00	44.00	0.00	0.00
	Black (n=50)	20.00*** ($h=-1.69$) [11.24, 33.04]	80.00*** ($h=1.69$) [66.96, 88.76]	18.00	62.00	0.00	0.00
	Hispanic (n=50)	0.00*** ($h=-2.47$) [0.00, 7.13]	100.00*** ($h=2.47$) [92.87, 100.00]	54.00	46.00	0.00	0.00
	Asian (n=50)	2.00*** ($h=-2.47$) [0.35, 10.50]	98.00*** ($h=2.47$) [89.50, 99.65]	30.00	68.00	0.00	0.00
Age	Baby Boomer (n=50)	0.00*** ($h=-2.43$) [0.00, 7.13] ^{†††}	100.00*** ($h=2.61$) [92.87, 100.00] ^{†††}	60.00	40.00	0.00	0.00
	Generation X (n=50)	0.00*** ($h=-2.29$) [0.00, 7.13] ^{†††}	100.00*** ($h=2.35$) [92.87, 100.00] ^{†††}	28.00	72.00	0.00	0.00
	Millennial (n=50)	0.00*** ($h=-2.05$) [0.00, 7.13]	100.00*** ($h=2.19$) [92.87, 100.00]	24.00	76.00	0.00	0.00
	Generation Z (n=50)	0.00*** ($h=-1.83$) [0.00, 7.13]	100.00*** ($h=1.94$) [92.87, 100.00]	4.00	96.00	0.00	0.00
	Generation Alpha (n=50)	10.00 ($h=N/A$) [4.35, 21.36] ^{†††}	90.00 ($h=N/A$) [78.64, 95.65] ^{†††}	10.00	80.00	0.00	0.00

Table A19. Sexual orientation analysis of implicit bias for llama-3.1-70b.

		command-r-plus					
		Heterosexual	LGBTQ	Homosexual	Bisexual	Other	Refusal
Gender	Male (n=500)	18.60*** ($h=-1.90$) [15.43, 22.25] ^{†††}	81.40*** ($h=1.90$) [77.75, 84.57] ^{†††}	77.40	3.20	0.80	0.00
	Female (n=500)	10.20*** ($h=-2.14$) [7.84, 13.16] ^{†††}	89.80*** ($h=2.21$) [86.84, 92.16] ^{†††}	81.20	7.20	1.40	0.00
Ethnicity/Race	Neutral (n=50)	4.00*** ($h=-2.18$) [1.10, 13.46] ^{†††}	96.00*** ($h=2.18$) [86.54, 98.90] ^{†††}	84.00	4.00	8.00	0.00
	White (n=50)	20.00*** ($h=-1.71$) [11.24, 33.04] ^{†††}	80.00*** ($h=1.71$) [66.96, 88.76] ^{†††}	64.00	8.00	8.00	0.00
	Black (n=50)	12.00*** ($h=-1.91$) [5.62, 23.80]	88.00*** ($h=1.91$) [76.20, 94.38]	70.00	12.00	6.00	0.00
	Hispanic (n=50)	2.00*** ($h=-2.18$) [0.35, 10.50] ^{†††}	98.00*** ($h=2.18$) [89.50, 99.65] ^{†††}	94.00	4.00	0.00	0.00
	Asian (n=50)	12.00*** ($h=-2.04$) [5.62, 23.80] [†]	88.00*** ($h=2.04$) [76.20, 94.38] [†]	84.00	4.00	0.00	0.00
Age	Baby Boomer (n=50)	14.00*** ($h=-1.67$) [6.95, 26.19] ^{†††}	86.00*** ($h=1.84$) [73.81, 93.05] ^{†††}	82.00	4.00	0.00	0.00
	Generation X (n=50)	8.00*** ($h=-1.72$) [3.15, 18.84] ^{†††}	92.00*** ($h=1.77$) [81.16, 96.85] ^{†††}	78.00	8.00	6.00	0.00
	Millennial (n=50)	6.00*** ($h=-1.55$) [2.06, 16.22] ^{†††}	94.00*** ($h=1.69$) [83.78, 97.94] ^{†††}	84.00	10.00	0.00	0.00
	Generation Z (n=50)	4.00*** ($h=-1.43$) [1.10, 13.46] ^{†††}	96.00*** ($h=1.54$) [86.54, 98.90] ^{†††}	90.00	4.00	2.00	0.00
	Generation Alpha (n=50)	8.00 ($h=N/A$) [3.15, 18.84] [†]	92.00 ($h=N/A$) [81.16, 96.85] [†]	74.00	12.00	6.00	0.00

Table A20. Sexual orientation analysis of implicit bias for command-r-plus.

Explicit

		claude-3.5-sonnet					
		Heterosexual	LGBTQ	Homosexual	Bisexual	Other	Refusal
Gender	Male (n=50)	0.00*** ($h=-2.79$) [0.00, 7.13]	100.00*** ($h=2.79$) [92.87, 100.00]	96.00	4.00	0.00	0.00
	Female (n=50)	0.00*** ($h=-2.79$) [0.00, 7.13] ^{†††}	100.00*** ($h=2.86$) [92.87, 100.00] ^{†††}	0.00	100.00	0.00	0.00
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-2.58$) [0.00, 7.13]	100.00*** ($h=2.58$) [92.87, 100.00]	46.00	54.00	0.00	0.00
	White (n=50)	0.00*** ($h=N/A$) [0.00, 0.00]	0.00*** ($h=N/A$) [0.00, 0.00]	0.00	0.00	0.00	100.00
	Black (n=50)	26.00*** ($h=-1.01$) [33.50, 69.97]	24.00*** ($h=1.01$) [30.03, 66.50]	0.00	24.00	0.00	50.00
	Hispanic (n=50)	76.00* ($h=-0.27$) [65.74, 88.27] ^{†††}	20.00* ($h=0.27$) [11.73, 34.26] ^{†††}	14.00	6.00	0.00	4.00
	Asian (n=50)	24.00*** ($h=-1.26$) [28.76, 64.54] [†]	28.00*** ($h=1.26$) [35.46, 71.24] [†]	26.00	2.00	0.00	48.00
Age	Baby Boomer (n=50)	100.00** ($h=0.71$) [92.87, 100.00] ^{†††}	0.00* ($h=-0.54$) [0.00, 7.13] ^{†††}	0.00	0.00	0.00	0.00
	Generation X (n=50)	74.00 ($h=-0.22$) [60.45, 84.13] ^{†††}	26.00* ($h=0.27$) [15.87, 39.55] ^{†††}	0.00	26.00	0.00	0.00
	Millennial (n=50)	8.00*** ($h=-1.48$) [3.15, 18.84]	92.00*** ($h=1.62$) [81.16, 96.85]	2.00	90.00	0.00	0.00
	Generation Z (n=50)	4.00*** ($h=-1.43$) [1.10, 13.46]	96.00*** ($h=1.54$) [86.54, 98.90]	0.00	96.00	0.00	0.00
	Generation Alpha (n=50)	4.00 ($h=N/A$) [1.10, 13.46]	96.00 ($h=N/A$) [86.54, 98.90]	0.00	96.00	0.00	0.00

Table A21. Sexual orientation analysis of explicit bias for claude-3.5-sonnet.

gpt-4o-mini							
		Heterosexual	LGBTQ	Homosexual	Bisexual	Other	Refusal
Gender	Male (n=50)	18.00*** ($h=-1.92$) [9.77, 30.80] [†]	82.00*** ($h=1.92$) [69.20, 90.23] [†]	60.00	22.00	0.00	0.00
	Female (n=50)	0.00*** ($h=-2.79$) [0.00, 7.13] ^{††}	100.00*** ($h=2.86$) [92.87, 100.00] ^{††}	0.00	100.00	0.00	0.00
Ethnicity/Race	Neutral (n=50)	14.00*** ($h=-1.82$) [6.95, 26.19]	86.00*** ($h=1.82$) [73.81, 93.05]	28.00	58.00	0.00	0.00
	White (n=50)	54.00*** ($h=-0.99$) [40.40, 67.03] ^{†††}	46.00*** ($h=0.99$) [32.97, 59.60] ^{†††}	4.00	42.00	0.00	0.00
	Black (n=50)	18.00*** ($h=-1.75$) [9.77, 30.80]	82.00*** ($h=1.75$) [69.20, 90.23]	8.00	74.00	0.00	0.00
	Hispanic (n=50)	26.00*** ($h=-1.40$) [15.87, 39.55] [†]	74.00*** ($h=1.40$) [60.45, 84.13] [†]	10.00	64.00	0.00	0.00
	Asian (n=50)	8.00*** ($h=-2.18$) [3.15, 18.84]	92.00*** ($h=2.18$) [81.16, 96.85]	36.00	56.00	0.00	0.00
Age	Baby Boomer (n=50)	98.00* ($h=0.42$) [89.50, 99.65] ^{†††}	2.00 ($h=-0.25$) [0.35, 10.50] ^{†††}	0.00	2.00	0.00	0.00
	Generation X (n=50)	48.00*** ($h=-0.76$) [34.80, 61.49] ^{†††}	52.00*** ($h=0.82$) [38.51, 65.20] ^{†††}	0.00	52.00	0.00	0.00
	Millennial (n=50)	0.00*** ($h=-2.05$) [0.00, 7.13]	100.00*** ($h=2.19$) [92.87, 100.00]	2.00	98.00	0.00	0.00
	Generation Z (n=50)	2.00*** ($h=-1.55$) [0.35, 10.50]	98.00*** ($h=1.66$) [89.50, 99.65]	0.00	98.00	0.00	0.00
	Generation Alpha (n=50)	22.00 ($h=N/A$) [12.75, 35.24]	78.00 ($h=N/A$) [64.76, 87.25]	0.00	78.00	0.00	0.00

Table A22. Sexual orientation analysis of explicit bias for gpt-4o-mini.

llama-3.1-70b							
		Heterosexual	LGBTQ	Homosexual	Bisexual	Other	Refusal
Gender	Male (n=50)	2.00*** ($h=-2.51$) [0.35, 10.50]	98.00*** ($h=2.51$) [89.50, 99.65]	96.00	2.00	0.00	0.00
	Female (n=50)	0.00*** ($h=-2.79$) [0.00, 7.13]	100.00*** ($h=2.86$) [92.87, 100.00]	6.00	94.00	0.00	0.00
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-2.58$) [0.00, 7.13]	100.00*** ($h=2.58$) [92.87, 100.00]	64.00	36.00	0.00	0.00
	White (n=50)	36.00*** ($h=-1.35$) [24.14, 49.86] ^{††}	64.00*** ($h=1.35$) [50.14, 75.86] ^{††}	32.00	32.00	0.00	0.00
	Black (n=50)	10.00*** ($h=-1.98$) [4.35, 21.36]	90.00*** ($h=1.98$) [78.64, 95.65]	68.00	22.00	0.00	0.00
	Hispanic (n=50)	4.00*** ($h=-2.06$) [1.10, 13.46]	96.00*** ($h=2.06$) [86.54, 98.90]	54.00	42.00	0.00	0.00
	Asian (n=50)	0.00*** ($h=-2.75$) [0.00, 7.13]	100.00*** ($h=2.75$) [92.87, 100.00]	58.00	42.00	0.00	0.00
Age	Baby Boomer (n=50)	100.00** ($h=0.71$) [92.87, 100.00] ^{†††}	0.00* ($h=-0.54$) [0.00, 7.13] ^{†††}	0.00	0.00	0.00	0.00
	Generation X (n=50)	68.00** ($h=-0.35$) [54.19, 79.24] ^{†††}	32.00** ($h=0.41$) [20.76, 45.81] ^{†††}	6.00	26.00	0.00	0.00
	Millennial (n=50)	10.00*** ($h=-1.41$) [4.35, 21.36]	90.00*** ($h=1.55$) [78.64, 95.65]	28.00	62.00	0.00	0.00
	Generation Z (n=50)	0.00*** ($h=-1.83$) [0.00, 7.13]	100.00*** ($h=1.94$) [92.87, 100.00]	16.00	84.00	0.00	0.00
	Generation Alpha (n=50)	80.00 ($h=N/A$) [66.96, 88.76] ^{†††}	20.00 ($h=N/A$) [11.24, 33.04] ^{†††}	6.00	14.00	0.00	0.00

Table A23. Sexual orientation analysis of explicit bias for llama-3.1-70b.

		command-r-plus					
		Heterosexual	LGBTQ	Homosexual	Bisexual	Other	Refusal
Gender	Male (n=50)	74.00*** ($h=-0.72$) [60.45, 84.13] ^{†††}	26.00*** ($h=0.72$) [15.87, 39.55] ^{†††}	24.00	2.00	0.00	0.00
	Female (n=50)	38.00*** ($h=-1.46$) [25.86, 51.85] ^{†††}	62.00*** ($h=1.53$) [48.15, 74.14] ^{†††}	58.00	4.00	0.00	0.00
Ethnicity/Race	Neutral (n=50)	40.00*** ($h=-1.21$) [27.61, 53.82] ^{†††}	60.00*** ($h=1.21$) [46.18, 72.39] ^{†††}	54.00	6.00	0.00	0.00
	White (n=50)	90.00 ($h=-0.14$) [78.64, 95.65] ^{†††}	10.00 ($h=0.14$) [4.35, 21.36] ^{†††}	6.00	4.00	0.00	0.00
	Black (n=50)	28.00*** ($h=-1.51$) [17.47, 41.67]	72.00*** ($h=1.51$) [58.33, 82.53]	72.00	0.00	0.00	0.00
	Hispanic (n=50)	50.00*** ($h=-0.89$) [36.64, 63.36] ^{†††}	50.00*** ($h=0.89$) [36.64, 63.36] ^{†††}	42.00	8.00	0.00	0.00
	Asian (n=50)	32.00*** ($h=-1.55$) [20.76, 45.81] [†]	68.00*** ($h=1.55$) [54.19, 79.24] [†]	56.00	12.00	0.00	0.00
Age	Baby Boomer (n=50)	100.00** ($h=0.71$) [92.87, 100.00] ^{†††}	0.00* ($h=-0.54$) [0.00, 7.13] ^{†††}	0.00	0.00	0.00	0.00
	Generation X (n=50)	94.00* ($h=0.36$) [83.78, 97.94] ^{†††}	6.00 ($h=-0.30$) [2.06, 16.22] ^{†††}	4.00	2.00	0.00	0.00
	Millennial (n=50)	52.00** ($h=-0.44$) [38.51, 65.20] ^{†††}	48.00*** ($h=0.58$) [34.80, 61.49] ^{†††}	28.00	20.00	0.00	0.00
	Generation Z (n=50)	34.00*** ($h=-0.59$) [22.44, 47.85] ^{†††}	66.00*** ($h=0.69$) [52.15, 77.56] ^{†††}	24.00	42.00	0.00	0.00
	Generation Alpha (n=50)	26.00 ($h=N/A$) [15.87, 39.55] [†]	74.00 ($h=N/A$) [60.45, 84.13] [†]	44.00	30.00	0.00	0.00

Table A24. Sexual orientation analysis of explicit bias for command-r-plus.

Socioeconomic Status Tables
Implicit

		claude-3.5-sonnet			
		Upper-class	Middle-class	Lower-class	Refusal
Gender	Male (n=500)	5.60*** ($h=-0.40$) [3.90, 7.97] ^{††}	93.80*** ($h=1.01$) [91.33, 95.60] ^{††}	0.60*** ($h=-0.96$) [0.20, 1.75]	0.00
	Female (n=500)	7.80*** ($h=-0.26$) [5.78, 10.53]	91.80*** ($h=0.98$) [89.47, 94.22]	0.00*** ($h=-1.20$) [0.00, 0.77]	0.40
Ethnicity/Race	Neutral (n=50)	8.00 ($h=-0.26$) [3.29, 19.55]	88.00*** ($h=0.95$) [80.45, 96.71]	0.00*** ($h=-1.16$) [0.00, 7.41]	4.00
	White (n=50)	38.00** ($h=0.43$) [27.64, 54.66]	52.00 ($h=0.01$) [41.25, 68.59]	4.00*** ($h=-0.61$) [1.17, 14.25]	6.00
	Black (n=50)	14.00 ($h=0.17$) [7.25, 27.17] ^{†††}	70.00*** ($h=0.56$) [59.00, 83.43] ^{†††}	12.00*** ($h=-0.75$) [5.86, 24.70]	4.00
	Hispanic (n=50)	28.00*** ($h=0.54$) [17.47, 41.67] ^{†††}	72.00** ($h=0.48$) [58.33, 82.53] ^{†††}	0.00*** ($h=-1.43$) [0.00, 7.13]	0.00
	Asian (n=50)	42.00** ($h=0.41$) [32.94, 60.92] ^{††}	48.00 ($h=0.11$) [39.08, 67.06] ^{††}	0.00*** ($h=-1.02$) [0.00, 7.87]	10.00
Age	Baby Boomer (n=50)	6.00 ($h=-0.30$) [2.06, 16.22] ^{†††}	94.00*** ($h=1.08$) [83.78, 97.94] ^{†††}	0.00*** ($h=-1.27$) [0.00, 7.13]	0.00
	Generation X (n=50)	2.00*** ($h=-0.74$) [0.35, 10.50]	98.00*** ($h=1.25$) [89.50, 99.65]	0.00*** ($h=-1.00$) [0.00, 7.13]	0.00
	Millennial (n=50)	2.00** ($h=-0.59$) [0.35, 10.50]	98.00*** ($h=1.19$) [89.50, 99.65]	0.00*** ($h=-1.07$) [0.00, 7.13]	0.00
	Generation Z (n=50)	8.00 ($h=-0.16$) [3.15, 18.84]	92.00*** ($h=0.88$) [81.16, 96.85]	0.00*** ($h=-1.18$) [0.00, 7.13]	0.00
	Generation Alpha (n=50)	22.00 ($h=0.24$) [12.75, 35.24] ^{†††}	78.00*** ($h=0.61$) [64.76, 87.25] ^{†††}	0.00*** ($h=-1.33$) [0.00, 7.13]	0.00

Table A25. Socioeconomic status analysis of implicit bias for claude-3.5-sonnet.

gpt-4o-mini					
		Upper-class	Middle-class	Lower-class	Refusal
Gender	Male (n=500)	0.00*** ($h=-0.88$) [0.00, 0.76]	70.60*** ($h=0.36$) [66.46, 74.42] [†]	29.40 ($h=0.03$) [25.58, 33.54] [†]	0.00
	Female (n=500)	0.00*** ($h=-0.82$) [0.00, 0.76]	77.00*** ($h=0.55$) [73.11, 80.47] ^{†††}	23.00*** ($h=-0.20$) [19.53, 26.89] ^{†††}	0.00
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-0.85$) [0.00, 7.13]	82.00*** ($h=0.65$) [69.20, 90.23]	18.00 ($h=-0.28$) [9.77, 30.80]	0.00
	White (n=50)	0.00*** ($h=-0.95$) [0.00, 7.13]	56.00 ($h=0.02$) [42.31, 68.84] ^{†††}	44.00** ($h=0.43$) [31.16, 57.69] ^{†††}	0.00
	Black (n=50)	6.00 ($h=-0.11$) [2.06, 16.22]	44.00 ($h=-0.04$) [31.16, 57.69] ^{†††}	50.00 ($h=0.10$) [36.64, 63.36] ^{†††}	0.00
	Hispanic (n=50)	0.00* ($h=-0.57$) [0.00, 7.13]	54.00 ($h=0.10$) [40.40, 67.03]	46.00 ($h=0.06$) [32.97, 59.60]	0.00
	Asian (n=50)	0.00*** ($h=-1.09$) [0.00, 7.13] [†]	72.00*** ($h=0.50$) [58.33, 82.53]	28.00 ($h=0.09$) [17.47, 41.67] ^{†††}	0.00
Age	Baby Boomer (n=50)	0.00*** ($h=-0.80$) [0.00, 7.13]	66.00* ($h=0.33$) [52.15, 77.56] ^{†††}	34.00 ($h=-0.02$) [22.44, 47.85] ^{†††}	0.00
	Generation X (n=50)	0.00*** ($h=-1.02$) [0.00, 7.13]	64.00 ($h=0.24$) [50.14, 75.86] ^{†††}	36.00* ($h=0.29$) [24.14, 49.86] ^{†††}	0.00
	Millennial (n=50)	0.00*** ($h=-0.88$) [0.00, 7.13]	78.00*** ($h=0.49$) [64.76, 87.25] ^{†††}	22.00 ($h=-0.09$) [12.75, 35.24] ^{†††}	0.00
	Generation Z (n=50)	0.00** ($h=-0.74$) [0.00, 7.13]	82.00*** ($h=0.57$) [69.20, 90.23] [†]	18.00* ($h=-0.30$) [9.77, 30.80] [†]	0.00
	Generation Alpha (n=50)	4.00 ($h=-0.34$) [1.10, 13.46]	58.00 ($h=0.18$) [44.23, 70.62] ^{†††}	38.00 ($h=0.00$) [25.86, 51.85] ^{†††}	0.00

Table A26. Socioeconomic status analysis of implicit bias for gpt-4o-mini.

llama-3.1-70b					
		Upper-class	Middle-class	Lower-class	Refusal
Gender	Male (n=500)	1.40*** ($h=-0.64$) [0.68, 2.86]	97.40*** ($h=1.19$) [95.60, 98.47]	1.20*** ($h=-0.90$) [0.55, 2.59]	0.00
	Female (n=500)	0.00*** ($h=-0.82$) [0.00, 0.76]	98.00*** ($h=1.27$) [96.36, 98.91]	2.00*** ($h=-0.92$) [1.09, 3.64]	0.00
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-0.85$) [0.00, 7.13]	100.00*** ($h=1.53$) [92.87, 100.00]	0.00*** ($h=-1.16$) [0.00, 7.13]	0.00
	White (n=50)	0.00*** ($h=-0.95$) [0.00, 7.13]	96.00*** ($h=1.07$) [86.54, 98.90]	4.00*** ($h=-0.62$) [1.10, 13.46]	0.00
	Black (n=50)	8.00 ($h=-0.04$) [3.15, 18.84]	76.00*** ($h=0.63$) [62.59, 85.70] ^{†††}	16.00*** ($h=-0.65$) [8.34, 28.51] ^{††}	0.00
	Hispanic (n=50)	0.00* ($h=-0.57$) [0.00, 7.13]	100.00*** ($h=1.59$) [92.87, 100.00]	0.00*** ($h=-1.43$) [0.00, 7.13]	0.00
	Asian (n=50)	0.00*** ($h=-1.09$) [0.00, 7.13]	96.00*** ($h=1.21$) [86.54, 98.90]	4.00*** ($h=-0.62$) [1.10, 13.46]	0.00
Age	Baby Boomer (n=50)	0.00*** ($h=-0.80$) [0.00, 7.13]	98.00*** ($h=1.29$) [89.50, 99.65]	2.00*** ($h=-0.98$) [0.35, 10.50]	0.00
	Generation X (n=50)	0.00*** ($h=-1.02$) [0.00, 7.13]	96.00*** ($h=1.13$) [86.54, 98.90]	4.00*** ($h=-0.60$) [1.10, 13.46]	0.00
	Millennial (n=50)	0.00*** ($h=-0.88$) [0.00, 7.13]	98.00*** ($h=1.19$) [89.50, 99.65]	2.00*** ($h=-0.79$) [0.35, 10.50]	0.00
	Generation Z (n=50)	2.00* ($h=-0.45$) [0.35, 10.50]	96.00*** ($h=1.05$) [86.54, 98.90]	2.00*** ($h=-0.90$) [0.35, 10.50]	0.00
	Generation Alpha (n=50)	4.00 ($h=-0.34$) [1.10, 13.46] [†]	92.00*** ($h=1.02$) [81.16, 96.85]	4.00*** ($h=-0.93$) [1.10, 13.46]	0.00

Table A27. Socioeconomic status analysis of implicit bias for llama-3.1-70b.

		command-r-plus			
		Upper-class	Middle-class	Lower-class	Refusal
Gender	Male (n=500)	42.20*** ($h=0.54$) [37.95, 46.57] ^{†††}	57.80* ($h=0.10$) [53.43, 62.05] ^{†††}	0.00*** ($h=-1.12$) [0.00, 0.76]	0.00
	Female (n=500)	31.80*** ($h=0.38$) [27.87, 36.01] ^{†††}	68.20*** ($h=0.35$) [63.99, 72.13] ^{†††}	0.00*** ($h=-1.20$) [0.00, 0.76]	0.00
Ethnicity/Race	Neutral (n=50)	16.00 ($h=-0.03$) [8.34, 28.51] ^{†††}	84.00*** ($h=0.71$) [71.49, 91.66] ^{†††}	0.00*** ($h=-1.16$) [0.00, 7.13]	0.00
	White (n=50)	20.00 ($h=-0.02$) [11.24, 33.04] ^{†††}	80.00*** ($h=0.54$) [66.96, 88.76] ^{†††}	0.00*** ($h=-1.02$) [0.00, 7.13]	0.00
	Black (n=50)	22.00** ($h=0.37$) [12.75, 35.24] ^{†††}	78.00*** ($h=0.67$) [64.76, 87.25] ^{†††}	0.00*** ($h=-1.47$) [0.00, 7.13]	0.00
	Hispanic (n=50)	16.00 ($h=0.25$) [8.34, 28.51]	84.00*** ($h=0.77$) [71.49, 91.66]	0.00*** ($h=-1.43$) [0.00, 7.13]	0.00
	Asian (n=50)	42.00* ($h=0.32$) [29.38, 55.77]	58.00 ($h=0.20$) [44.23, 70.62]	0.00*** ($h=-1.02$) [0.00, 7.13]	0.00
Age	Baby Boomer (n=50)	24.00 ($h=0.23$) [14.30, 37.41]	76.00*** ($h=0.55$) [62.59, 85.70]	0.00*** ($h=-1.27$) [0.00, 7.13]	0.00
	Generation X (n=50)	22.00 ($h=-0.05$) [12.75, 35.24]	78.00*** ($h=0.55$) [64.76, 87.25]	0.00*** ($h=-1.00$) [0.00, 7.13]	0.00
	Millennial (n=50)	44.00*** ($h=0.57$) [31.16, 57.69] ^{††}	56.00 ($h=0.02$) [42.31, 68.84] ^{††}	0.00*** ($h=-1.07$) [0.00, 7.13]	0.00
	Generation Z (n=50)	30.00** ($h=0.42$) [19.10, 43.75] ^{†††}	70.00* ($h=0.29$) [56.25, 80.90] ^{†††}	0.00*** ($h=-1.18$) [0.00, 7.13]	0.00
	Generation Alpha (n=50)	36.00*** ($h=0.55$) [24.14, 49.86]	62.00 ($h=0.26$) [48.15, 74.14]	2.00*** ($h=-1.04$) [0.35, 10.50]	0.00

Table A28. Socioeconomic status analysis of implicit bias for command-r-plus.

Explicit

		claude-3.5-sonnet			
		Upper-class	Middle-class	Lower-class	Refusal
Gender	Male (n=50)	18.00 ($h=0.00$) [9.77, 30.80] ^{††}	82.00*** ($h=0.63$) [69.20, 90.23] ^{††}	0.00*** ($h=-1.12$) [0.00, 7.13]	0.00
	Female (n=50)	6.00 ($h=-0.33$) [2.06, 16.22]	94.00*** ($h=1.06$) [83.78, 97.94]	0.00*** ($h=-1.20$) [0.00, 7.13]	0.00
Ethnicity/Race	Neutral (n=50)	8.00 ($h=-0.28$) [3.15, 18.84]	92.00*** ($h=0.96$) [81.16, 96.85]	0.00*** ($h=-1.16$) [0.00, 7.13]	0.00
	White (n=50)	0.00*** ($h=N/A$) [0.00, 0.00]	0.00*** ($h=N/A$) [0.00, 0.00]	0.00*** ($h=N/A$) [0.00, 0.00]	100.00
	Black (n=50)	42.00*** ($h=1.71$) [65.35, 93.60] ^{†††}	8.00** ($h=-0.67$) [6.40, 34.65] ^{†††}	0.00*** ($h=-1.47$) [0.00, 13.32]	50.00
	Hispanic (n=50)	0.00* ($h=-0.57$) [0.00, 7.41] ^{†††}	96.00*** ($h=1.59$) [92.59, 100.00] ^{†††}	0.00*** ($h=-1.43$) [0.00, 7.41]	4.00
	Asian (n=50)	8.00 ($h=-0.29$) [6.15, 33.53] ^{††}	44.00*** ($h=0.80$) [66.47, 93.85] ^{††}	0.00** ($h=-1.02$) [0.00, 12.87]	48.00
Age	Baby Boomer (n=50)	38.00*** ($h=0.53$) [25.86, 51.85] ^{†††}	62.00 ($h=0.24$) [48.15, 74.14] ^{†††}	0.00*** ($h=-1.27$) [0.00, 7.13]	0.00
	Generation X (n=50)	12.00* ($h=-0.32$) [5.62, 23.80]	88.00*** ($h=0.82$) [76.20, 94.38]	0.00*** ($h=-1.00$) [0.00, 7.13]	0.00
	Millennial (n=50)	0.00*** ($h=-0.88$) [0.00, 7.13]	100.00*** ($h=1.47$) [92.87, 100.00]	0.00*** ($h=-1.07$) [0.00, 7.13]	0.00
	Generation Z (n=50)	0.00** ($h=-0.74$) [0.00, 7.13]	100.00*** ($h=1.45$) [92.87, 100.00]	0.00*** ($h=-1.18$) [0.00, 7.13]	0.00
	Generation Alpha (n=50)	86.00*** ($h=1.64$) [73.81, 93.05] ^{†††}	14.00*** ($h=-0.78$) [6.95, 26.19] ^{†††}	0.00*** ($h=-1.33$) [0.00, 7.13]	0.00

Table A29. Socioeconomic status analysis of explicit bias for claude-3.5-sonnet.

gpt-4o-mini					
		Upper-class	Middle-class	Lower-class	Refusal
Gender	Male (n=50)	2.00** ($h=-0.59$) [0.35, 10.50]	84.00*** ($h=0.69$) [71.49, 91.66] [†]	14.00* ($h=-0.35$) [6.95, 26.19] [†]	0.00
	Female (n=50)	0.00*** ($h=-0.82$) [0.00, 7.13]	96.00*** ($h=1.15$) [86.54, 98.90] ^{†††}	4.00*** ($h=-0.80$) [1.10, 13.46] ^{†††}	0.00
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-0.85$) [0.00, 7.13]	94.00*** ($h=1.04$) [83.78, 97.94]	6.00*** ($h=-0.66$) [2.06, 16.22]	0.00
	White (n=50)	6.00** ($h=-0.46$) [2.06, 16.22]	94.00*** ($h=0.98$) [83.78, 97.94] ^{†††}	0.00*** ($h=-1.02$) [0.00, 7.13] ^{†††}	0.00
	Black (n=50)	0.00* ($h=-0.61$) [0.00, 7.13]	86.00*** ($h=0.88$) [73.81, 93.05] ^{†††}	14.00*** ($h=-0.70$) [6.95, 26.19] ^{†††}	0.00
	Hispanic (n=50)	0.00* ($h=-0.57$) [0.00, 7.13]	62.00 ($h=0.26$) [48.15, 74.14]	38.00 ($h=-0.10$) [25.86, 51.85]	0.00
	Asian (n=50)	12.00* ($h=-0.39$) [5.62, 23.80] [†]	88.00*** ($h=0.90$) [76.20, 94.38]	0.00*** ($h=-1.02$) [0.00, 7.13] ^{†††}	0.00
Age	Baby Boomer (n=50)	0.00*** ($h=-0.80$) [0.00, 7.13]	100.00*** ($h=1.57$) [92.87, 100.00] ^{†††}	0.00*** ($h=-1.27$) [0.00, 7.13] ^{†††}	0.00
	Generation X (n=50)	0.00*** ($h=-1.02$) [0.00, 7.13]	100.00*** ($h=1.53$) [92.87, 100.00] ^{†††}	0.00*** ($h=-1.00$) [0.00, 7.13] ^{†††}	0.00
	Millennial (n=50)	0.00*** ($h=-0.88$) [0.00, 7.13]	100.00*** ($h=1.47$) [92.87, 100.00] ^{†††}	0.00*** ($h=-1.07$) [0.00, 7.13] ^{†††}	0.00
	Generation Z (n=50)	0.00** ($h=-0.74$) [0.00, 7.13]	98.00*** ($h=1.17$) [89.50, 99.65] [†]	2.00*** ($h=-0.90$) [0.35, 10.50] [†]	0.00
	Generation Alpha (n=50)	4.00 ($h=-0.34$) [1.10, 13.46]	94.00*** ($h=1.10$) [83.78, 97.94] ^{†††}	2.00*** ($h=-1.04$) [0.35, 10.50] ^{†††}	0.00

Table A30. Socioeconomic status analysis of explicit bias for gpt-4o-mini.

llama-3.1-70b					
		Upper-class	Middle-class	Lower-class	Refusal
Gender	Male (n=50)	2.00** ($h=-0.59$) [0.35, 10.50]	98.00*** ($h=1.23$) [89.50, 99.65]	0.00*** ($h=-1.12$) [0.00, 7.13]	0.00
	Female (n=50)	0.00*** ($h=-0.82$) [0.00, 7.13]	100.00*** ($h=1.55$) [92.87, 100.00]	0.00*** ($h=-1.20$) [0.00, 7.13]	0.00
Ethnicity/Race	Neutral (n=50)	0.00*** ($h=-0.85$) [0.00, 7.13]	100.00*** ($h=1.53$) [92.87, 100.00]	0.00*** ($h=-1.16$) [0.00, 7.13]	0.00
	White (n=50)	0.00*** ($h=-0.95$) [0.00, 7.13]	100.00*** ($h=1.47$) [92.87, 100.00]	0.00*** ($h=-1.02$) [0.00, 7.13]	0.00
	Black (n=50)	0.00* ($h=-0.61$) [0.00, 7.13]	100.00*** ($h=1.65$) [92.87, 100.00] ^{†††}	0.00*** ($h=-1.47$) [0.00, 7.13] ^{††}	0.00
	Hispanic (n=50)	0.00* ($h=-0.57$) [0.00, 7.13]	100.00*** ($h=1.59$) [92.87, 100.00]	0.00*** ($h=-1.43$) [0.00, 7.13]	0.00
	Asian (n=50)	10.00** ($h=-0.45$) [4.35, 21.36]	90.00*** ($h=0.97$) [78.64, 95.65]	0.00*** ($h=-1.02$) [0.00, 7.13]	0.00
Age	Baby Boomer (n=50)	0.00*** ($h=-0.80$) [0.00, 7.13]	100.00*** ($h=1.57$) [92.87, 100.00]	0.00*** ($h=-1.27$) [0.00, 7.13]	0.00
	Generation X (n=50)	2.00*** ($h=-0.74$) [0.35, 10.50]	98.00*** ($h=1.25$) [89.50, 99.65]	0.00*** ($h=-1.00$) [0.00, 7.13]	0.00
	Millennial (n=50)	4.00** ($h=-0.47$) [1.10, 13.46]	96.00*** ($h=1.07$) [86.54, 98.90]	0.00*** ($h=-1.07$) [0.00, 7.13]	0.00
	Generation Z (n=50)	0.00** ($h=-0.74$) [0.00, 7.13]	100.00*** ($h=1.45$) [92.87, 100.00]	0.00*** ($h=-1.18$) [0.00, 7.13]	0.00
	Generation Alpha (n=50)	22.00 ($h=0.24$) [12.75, 35.24] [†]	78.00*** ($h=0.61$) [64.76, 87.25]	0.00*** ($h=-1.33$) [0.00, 7.13]	0.00

Table A31. Socioeconomic status analysis of explicit bias for llama-3.1-70b.

command-r-plus					
		Upper-class	Middle-class	Lower-class	Refusal
Gender	Male (n=50)	86.00*** ($h=1.50$) [73.81, 93.05] ^{†††}	14.00*** ($h=-0.86$) [6.95, 26.19] ^{†††}	0.00*** ($h=-1.12$) [0.00, 7.13]	0.00
	Female (n=50)	92.00*** ($h=1.75$) [81.16, 96.85] ^{†††}	6.00*** ($h=-1.10$) [2.06, 16.22] ^{†††}	2.00*** ($h=-0.92$) [0.35, 10.50]	0.00
Ethnicity/Race	Neutral (n=50)	82.00*** ($h=1.42$) [69.20, 90.23] ^{†††}	16.00*** ($h=-0.79$) [8.34, 28.51] ^{†††}	2.00*** ($h=-0.88$) [0.35, 10.50]	0.00
	White (n=50)	90.00*** ($h=1.55$) [78.64, 95.65] ^{†††}	10.00*** ($h=-1.03$) [4.35, 21.36] ^{†††}	0.00*** ($h=-1.02$) [0.00, 7.13]	0.00
	Black (n=50)	60.00*** ($h=1.16$) [46.18, 72.39] ^{†††}	40.00 ($h=-0.12$) [27.61, 53.82] ^{†††}	0.00*** ($h=-1.47$) [0.00, 7.13]	0.00
	Hispanic (n=50)	24.00*** ($h=0.45$) [14.30, 37.41]	76.00*** ($h=0.57$) [62.59, 85.70]	0.00*** ($h=-1.43$) [0.00, 7.13]	0.00
	Asian (n=50)	62.00*** ($h=0.72$) [48.15, 74.14]	38.00 ($h=-0.20$) [25.86, 51.85]	0.00*** ($h=-1.02$) [0.00, 7.13]	0.00
Age	Baby Boomer (n=50)	30.00** ($h=0.36$) [19.10, 43.75]	68.00* ($h=0.37$) [54.19, 79.24]	2.00*** ($h=-0.98$) [0.35, 10.50]	0.00
	Generation X (n=50)	20.00 ($h=-0.10$) [11.24, 33.04]	78.00*** ($h=0.55$) [64.76, 87.25]	2.00*** ($h=-0.72$) [0.35, 10.50]	0.00
	Millennial (n=50)	16.00 ($h=-0.05$) [8.34, 28.51] ^{††}	84.00*** ($h=0.65$) [71.49, 91.66] ^{††}	0.00*** ($h=-1.07$) [0.00, 7.13]	0.00
	Generation Z (n=50)	4.00 ($h=-0.34$) [1.10, 13.46] ^{†††}	96.00*** ($h=1.05$) [86.54, 98.90] ^{†††}	0.00*** ($h=-1.18$) [0.00, 7.13]	0.00
	Generation Alpha (n=50)	24.00* ($h=0.29$) [14.30, 37.41]	76.00*** ($h=0.57$) [62.59, 85.70]	0.00*** ($h=-1.33$) [0.00, 7.13]	0.00

Table A32. Socioeconomic status analysis of explicit bias for command-r-plus.

Occupation Tables

Implicit

claude-3.5-sonnet		
Most Popular Occupations		
Gender	Male	engineer (95.2%), accountant (2.2%), mechanic (1.2%), teacher (0.8%), electrician (0.6%)
	Female	teacher (24.4%), engineer (22.2%), executive (21.4%), nurse (15.4%), veterinarian (13.2%)
Ethnicity/Race	Neutral	engineer (68.0%), nurse (8.0%), executive (8.0%), biologist (6.0%), refusal (4.0%)
	White	engineer (78.0%), designer (8.0%), refusal (6.0%), teacher (4.0%), developer (2.0%)
	Black	engineer (50.0%), designer (16.0%), nurse (10.0%), driver (6.0%), teacher (6.0%)
	Hispanic	engineer (86.0%), biologist (2.0%), teacher (2.0%), veterinarian (2.0%), designer (2.0%)
	Asian	engineer (64.0%), executive (20.0%), refusal (10.0%), lawyer (2.0%), teacher (2.0%)
Age	Baby Boomer	teacher (40.0%), engineer (38.0%), veterinarian (8.0%), accountant (8.0%), librarian (2.0%)
	Generation X	engineer (54.0%), executive (16.0%), teacher (12.0%), veterinarian (12.0%), nurse (6.0%)
	Millennial	engineer (56.0%), executive (22.0%), veterinarian (14.0%), designer (4.0%), teacher (2.0%)
	Generation Z	engineer (74.0%), executive (14.0%), nurse (8.0%), designer (2.0%), veterinarian (2.0%)
	Generation Alpha	engineer (58.0%), president (10.0%), executive (10.0%), biologist (10.0%), designer (8.0%)

Table A33. Occupation analysis of implicit bias for claude-3.5-sonnet.

gpt-4o-mini		
Most Popular Occupations		
Gender	Male	graphic designer (47.6%), teacher (21.6%), social worker (13.8%), community organizer (5.4%), software engineer (4.4%)
	Female	graphic designer (33.6%), social worker (33.0%), teacher (18.8%), nurse (5.6%), community organizer (4.2%)
Ethnicity/Race	Neutral	graphic designer (48.0%), social worker (16.0%), teacher (16.0%), environmental scientist (8.0%), software developer (6.0%)
	White	social worker (44.0%), community organizer (16.0%), graphic designer (14.0%), teacher (10.0%), organizer (8.0%)
	Black	community organizer (26.0%), social worker (22.0%), graphic designer (22.0%), designer (8.0%), community health worker (4.0%)
	Hispanic	graphic designer (28.0%), community organizer (20.0%), social worker (20.0%), teacher (14.0%), organizer (4.0%)
	Asian	social worker (26.0%), teacher (22.0%), community organizer (18.0%), graphic designer (12.0%), software engineer (6.0%)
Age	Baby Boomer	teacher (44.0%), graphic designer (22.0%), social worker (20.0%), community organizer (6.0%), nurse (6.0%)
	Generation X	graphic designer (50.0%), social worker (20.0%), teacher (18.0%), nurse (8.0%), software engineer (2.0%)
	Millennial	graphic designer (58.0%), social worker (16.0%), teacher (12.0%), community organizer (4.0%), environmental scientist (4.0%)
	Generation Z	graphic designer (62.0%), social worker (16.0%), teacher (8.0%), designer (6.0%), environmental scientist (4.0%)
	Generation Alpha	graphic designer (50.0%), social worker (14.0%), teacher (10.0%), designer (6.0%), president (6.0%)

Table A34. Occupation analysis of implicit bias for gpt-4o-mini.

llama-3.1-70b		
Most Popular Occupations		
Gender	Male	teacher (54.6%), software engineer (19.8%), graphic designer (16.8%), writer (4.0%), marketing specialist (1.4%)
	Female	teacher (52.2%), graphic designer (18.4%), marketing specialist (14.0%), social worker (5.0%), writer (2.6%)
Ethnicity/Race	Neutral	graphic designer (38.0%), teacher (34.0%), software engineer (20.0%), environmental scientist (2.0%), writer (2.0%)
	White	rabbi (24.0%), social worker (24.0%), writer (20.0%), software engineer (14.0%), teacher (6.0%)
	Black	writer (22.0%), graphic designer (22.0%), midwife (10.0%), software engineer (8.0%), community organizer (8.0%)
	Hispanic	teacher (50.0%), graphic designer (42.0%), writer (4.0%), urban planner (2.0%), event planner (2.0%)
	Asian	graphic designer (26.0%), monk (20.0%), journalist (14.0%), software engineer (14.0%), writer (8.0%)
Age	Baby Boomer	teacher (70.0%), software engineer (6.0%), writer (6.0%), social worker (6.0%), graphic designer (4.0%)
	Generation X	graphic designer (54.0%), teacher (22.0%), marketing specialist (10.0%), social worker (6.0%), writer (4.0%)
	Millennial	graphic designer (54.0%), marketing specialist (16.0%), software engineer (14.0%), teacher (8.0%), writer (4.0%)
	Generation Z	graphic designer (76.0%), teacher (10.0%), software engineer (8.0%), marketing specialist (2.0%), writer (2.0%)
	Generation Alpha	graphic designer (58.0%), writer (20.0%), lawyer/politician (6.0%), event planner (6.0%), lawyer (4.0%)

Table A35. Occupation analysis of implicit bias for llama-3.1-70b.

command-r-plus		
Most Popular Occupations		
Gender	Male	financial analyst (31.6%), teacher (18.8%), freelance graphic designer (12.6%), artist (8.8%), software engineer (8.6%)
	Female	social worker (26.8%), artist (14.2%), freelance graphic designer (8.4%), teacher (7.6%), lawyer (6.2%)
Ethnicity/Race	Neutral	freelance graphic designer (18.0%), social worker (18.0%), artist (12.0%), teacher (12.0%), software engineer (10.0%)
	White	social worker (20.0%), rabbi (16.0%), teacher (10.0%), lawyer (8.0%), freelance graphic designer (6.0%)
	Black	social worker (14.0%), artist (10.0%), freelance graphic designer (8.0%), entrepreneur (6.0%), photographer (6.0%)
	Hispanic	teacher (32.0%), financial analyst (10.0%), investment banker (10.0%), social worker (10.0%), freelance graphic designer (6.0%)
	Asian	lawyer (20.0%), entrepreneur (12.0%), social worker (12.0%), artist (12.0%), software engineer (8.0%)
Age	Baby Boomer	teacher (30.0%), social worker (26.0%), artist (12.0%), financial analyst (10.0%), freelance graphic designer (4.0%)
	Generation X	social worker (22.0%), financial analyst (14.0%), entrepreneur (10.0%), artist (10.0%), attorney (8.0%)
	Millennial	freelance graphic designer (16.0%), social worker (12.0%), artist (12.0%), software engineer (10.0%), entrepreneur (10.0%)
	Generation Z	freelance graphic designer (14.0%), social worker (14.0%), lawyer (14.0%), artist (10.0%), finance (6.0%)
	Generation Alpha	artist (36.0%), freelance graphic designer (14.0%), photographer (6.0%), attorney (6.0%), lawyer (6.0%)

Table A36. Occupation analysis of implicit bias for command-r-plus.

Explicit

claude-3.5-sonnet		
Most Popular Occupations		
Gender	Male	engineer (100.0%)
	Female	engineer (80.0%), veterinarian (14.0%), lawyer (6.0%)
Ethnicity/Race	Neutral	engineer (86.0%), veterinarian (10.0%), lawyer (4.0%)
	White	refusal (100.0%)
	Black	refusal (50.0%), neurosurgeon (40.0%), engineer (10.0%)
	Hispanic	engineer (44.0%), teacher (24.0%), nurse (20.0%), executive (8.0%), refusal (4.0%)
Age	Asian	engineer (52.0%), refusal (48.0%)
	Baby Boomer	teacher (46.0%), engineer (44.0%), accountant (4.0%), nurse (4.0%), manager (2.0%)
	Generation X	developer (42.0%), manager (34.0%), executive (10.0%), engineer (8.0%), director (6.0%)
	Millennial	specialist (52.0%), developer (46.0%), marketer (2.0%)
	Generation Z	designer (48.0%), developer (44.0%), specialist (4.0%), creator (2.0%), manager (2.0%)
	Generation Alpha	designer (56.0%), creator (34.0%), student (8.0%), youtuber (2.0%)

Table A37. Occupation analysis of explicit bias for claude-3.5-sonnet.

gpt-4o-mini		
Most Popular Occupations		
Gender	Male	teacher (54.0%), software engineer (24.0%), graphic designer (18.0%), social worker (4.0%)
	Female	environmental scientist (44.0%), graphic designer (24.0%), social worker (16.0%), teacher (4.0%), community organizer (4.0%)
Ethnicity/Race	Neutral	teacher (42.0%), environmental scientist (24.0%), social worker (12.0%), graphic designer (6.0%), software developer (6.0%)
	White	software engineer (30.0%), marketing manager (28.0%), teacher (26.0%), project manager (8.0%), graphic designer (8.0%)
	Black	community organizer (64.0%), social worker (22.0%), teacher (10.0%), community outreach coordinator (4.0%)
	Hispanic	community organizer (48.0%), community health worker (20.0%), construction foreman (10.0%), social worker (8.0%), mechanic (4.0%)
	Asian	software engineer (92.0%), graphic designer (4.0%), social worker (2.0%), software developer (2.0%)
Age	Baby Boomer	teacher (90.0%), retired school principal (4.0%), retired teacher (4.0%), retired schoolteacher (2.0%)
	Generation X	marketing manager (38.0%), project manager (34.0%), graphic designer (16.0%), software developer (6.0%), software engineer (2.0%)
	Millennial	digital marketing specialist (72.0%), software developer (6.0%), graphic designer (6.0%), digital marketing manager (6.0%), marketing manager (6.0%)
	Generation Z	social media manager (58.0%), digital marketing specialist (20.0%), graphic designer (12.0%), freelance graphic designer (6.0%), sustainability consultant (4.0%)
	Generation Alpha	student (38.0%), software developer (16.0%), digital content creator (12.0%), digital marketing specialist (8.0%), content creator (4.0%)

Table A38. Occupation analysis of explicit bias for gpt-4o-mini.

llama-3.1-70b		
Most Popular Occupations		
Gender	Male	software engineer (44.0%), teacher (42.0%), financial analyst (6.0%), environmental engineer (4.0%), data analyst (2.0%)
	Female	teacher (54.0%), dentist (10.0%), software engineer (10.0%), environmental scientist (10.0%), marketing specialist (10.0%)
Ethnicity/Race	Neutral	teacher (42.0%), software engineer (26.0%), environmental scientist (10.0%), financial analyst (6.0%), dentist (6.0%)
	White	software engineer (48.0%), marketing specialist (34.0%), marketing manager (16.0%), teacher (2.0%)
	Black	teacher (74.0%), principal (10.0%), marketing specialist (4.0%), dentist (4.0%), psychologist (4.0%)
	Hispanic	teacher (74.0%), nurse (24.0%), registered nurse (2.0%)
	Asian	dentist (46.0%), software engineer (42.0%), cardiologist (8.0%), pediatrician (2.0%), dermatologist (2.0%)
Age	Baby Boomer	teacher (38.0%), accountant (26.0%), librarian (14.0%), manager (8.0%), mechanic (6.0%)
	Generation X	software engineer (44.0%), marketing specialist (42.0%), teacher (4.0%), marketing manager (4.0%), electrician (2.0%)
	Millennial	software engineer (50.0%), marketing specialist (50.0%)
	Generation Z	graphic designer (36.0%), social media influencer (34.0%), software engineer (18.0%), social media manager (10.0%), digital content creator (2.0%)
	Generation Alpha	student (68.0%), robotics engineer (10.0%), software engineer (8.0%), environmental scientist (4.0%), climate scientist (4.0%)

Table A39. Occupation analysis of explicit bias for llama-3.1-70b.

command-r-plus		
Most Popular Occupations		
Gender	Male	lawyer (30.0%), finance (16.0%), financial analyst (12.0%), financier (8.0%), businessman (4.0%)
	Female	ceo (18.0%), prima ballerina (10.0%), executive (10.0%), lawyer (8.0%), corporate lawyer (6.0%)
Ethnicity/Race	Neutral	lawyer (22.0%), ceo (16.0%), financial analyst (10.0%), software engineer (6.0%), journalist (4.0%)
	White	lawyer (20.0%), businessman (14.0%), financial analyst (12.0%), stockbroker (8.0%), fashion designer (6.0%)
	Black	lawyer (36.0%), financial analyst (22.0%), attorney (8.0%), teacher (6.0%), judge (6.0%)
	Hispanic	lawyer (34.0%), teacher (12.0%), entrepreneur (10.0%), social worker (8.0%), mechanic (6.0%)
	Asian	financial analyst (24.0%), software engineer (22.0%), entrepreneur (10.0%), surgeon (8.0%), businessman (4.0%)
Age	Baby Boomer	teacher (26.0%), businessman (10.0%), retired school teacher (8.0%), manager (8.0%), executive (6.0%)
	Generation X	freelance graphic designer (12.0%), contractor (12.0%), marketing director (10.0%), self-employed graphic designer (8.0%), teacher (6.0%)
	Millennial	freelance graphic designer (36.0%), financial analyst (16.0%), marketing specialist (12.0%), entrepreneur (6.0%), social media manager (4.0%)
	Generation Z	social media manager (22.0%), social media influencer (18.0%), content creator (8.0%), freelance content creator (6.0%), entrepreneur (6.0%)
	Generation Alpha	entrepreneur (20.0%), social media influencer (16.0%), influencer (12.0%), software engineer (8.0%), content creator (6.0%)

Table A40. Occupation analysis of explicit bias for command-r-plus.

Polarity Tables

Implicit

claude-3.5-sonnet				
		Median	Standard Deviation	Refusal
Gender	Male	0.12	0.04	0.00
	Female	0.12	0.05	0.40
Ethnicity/Race	Neutral	0.12	0.06	4.00
	White	0.13	0.04	6.00
	Black	0.13	0.04	4.00
	Hispanic	0.11	0.04	0.00
	Asian	0.12	0.06	10.00
Age	Baby Boomer	0.14	0.05	0.00
	Generation X	0.12	0.04	0.00
	Millennial	0.10	0.05	0.00
	Generation Z	0.11	0.05	0.00
	Generation Alpha	0.13	0.04	0.00

Table A41. Polarity analysis of implicit bias for claude-3.5-sonnet.

gpt-4o-mini				
		Median	Standard Deviation	Refusal
Gender	Male	0.11	0.04	0.00
	Female	0.14	0.04	0.00
Ethnicity/Race	Neutral	0.14	0.05	0.00
	White	0.13	0.05	0.00
	Black	0.15	0.06	0.00
	Hispanic	0.12	0.04	0.00
	Asian	0.12	0.05	0.00
Age	Baby Boomer	0.13	0.06	0.00
	Generation X	0.13	0.05	0.00
	Millennial	0.14	0.04	0.00
	Generation Z	0.14	0.04	0.00
	Generation Alpha	0.14	0.04	0.00

Table A42. Polarity analysis of implicit bias for gpt-4o-mini.

llama-3.1-70b				
		Median	Standard Deviation	Refusal
Gender	Male	0.14	0.06	0.00
	Female	0.15	0.05	0.00
Ethnicity/Race	Neutral	0.17	0.05	0.00
	White	0.18	0.07	0.00
	Black	0.17	0.07	0.00
	Hispanic	0.17	0.06	0.00
	Asian	0.13	0.07	0.00
Age	Baby Boomer	0.15	0.06	0.00
	Generation X	0.14	0.06	0.00
	Millennial	0.16	0.06	0.00
	Generation Z	0.17	0.06	0.00
	Generation Alpha	0.16	0.06	0.00

Table A43. Polarity analysis of implicit bias for llama-3.1-70b.

command-r-plus				
		Median	Standard Deviation	Refusal
Gender	Male	0.12	0.06	0.00
	Female	0.15	0.06	0.00
Ethnicity/Race	Neutral	0.14	0.06	0.00
	White	0.15	0.07	0.00
	Black	0.17	0.08	0.00
	Hispanic	0.14	0.06	0.00
	Asian	0.16	0.06	0.00
Age	Baby Boomer	0.11	0.06	0.00
	Generation X	0.14	0.05	0.00
	Millennial	0.17	0.07	0.00
	Generation Z	0.14	0.05	0.00
	Generation Alpha	0.15	0.07	0.00

Table A44. Polarity analysis of implicit bias for command-r-plus.

claude-3.5-sonnet				
		Median	Standard Deviation	Refusal
Gender	Male	0.10	0.04	0.00
	Female	0.10	0.03	0.00
Ethnicity/Race	Neutral	0.11	0.04	0.00
	White	0.00	0.00	100.00
	Black	0.21	0.04	50.00
	Hispanic	0.12	0.05	4.00
	Asian	0.10	0.05	48.00
Age	Baby Boomer	0.11	0.04	0.00
	Generation X	0.06	0.05	0.00
	Millennial	0.06	0.03	0.00
	Generation Z	0.07	0.04	0.00
	Generation Alpha	0.08	0.04	22.00

Table A45. Polarity analysis of explicit bias for claude-3.5-sonnet.

gpt-4o-mini				
		Median	Standard Deviation	Refusal
Gender	Male	0.09	0.04	0.00
	Female	0.13	0.05	0.00
Ethnicity/Race	Neutral	0.11	0.05	0.00
	White	0.08	0.05	0.00
	Black	0.14	0.06	0.00
	Hispanic	0.11	0.04	0.00
	Asian	0.11	0.05	0.00
Age	Baby Boomer	0.15	0.05	0.00
	Generation X	0.08	0.06	0.00
	Millennial	0.13	0.05	0.00
	Generation Z	0.13	0.05	0.00
	Generation Alpha	0.14	0.06	0.00

Table A46. Polarity analysis of explicit bias for gpt-4o-mini.

llama-3.1-70b				
		Median	Standard Deviation	Refusal
Gender	Male	0.11	0.06	0.00
	Female	0.11	0.05	0.00
Ethnicity/Race	Neutral	0.14	0.05	0.00
	White	0.10	0.05	0.00
	Black	0.18	0.06	0.00
	Hispanic	0.17	0.04	0.00
	Asian	0.14	0.06	0.00
Age	Baby Boomer	0.16	0.05	0.00
	Generation X	0.08	0.06	0.00
	Millennial	0.16	0.04	0.00
	Generation Z	0.14	0.05	0.00
	Generation Alpha	0.16	0.05	0.00

Table A47. Polarity analysis of explicit bias for llama-3.1-70b.

command-r-plus				
		Median	Standard Deviation	Refusal
Gender	Male	0.18	0.06	0.00
	Female	0.22	0.07	0.00
Ethnicity/Race	Neutral	0.18	0.07	0.00
	White	0.16	0.07	0.00
	Black	0.18	0.07	0.00
	Hispanic	0.19	0.06	0.00
	Asian	0.10	0.07	0.00
Age	Baby Boomer	0.13	0.05	0.00
	Generation X	0.12	0.06	0.00
	Millennial	0.09	0.06	0.00
	Generation Z	0.12	0.06	0.00
	Generation Alpha	0.12	0.05	0.00

Table A48. Polarity analysis of explicit bias for command-r-plus.