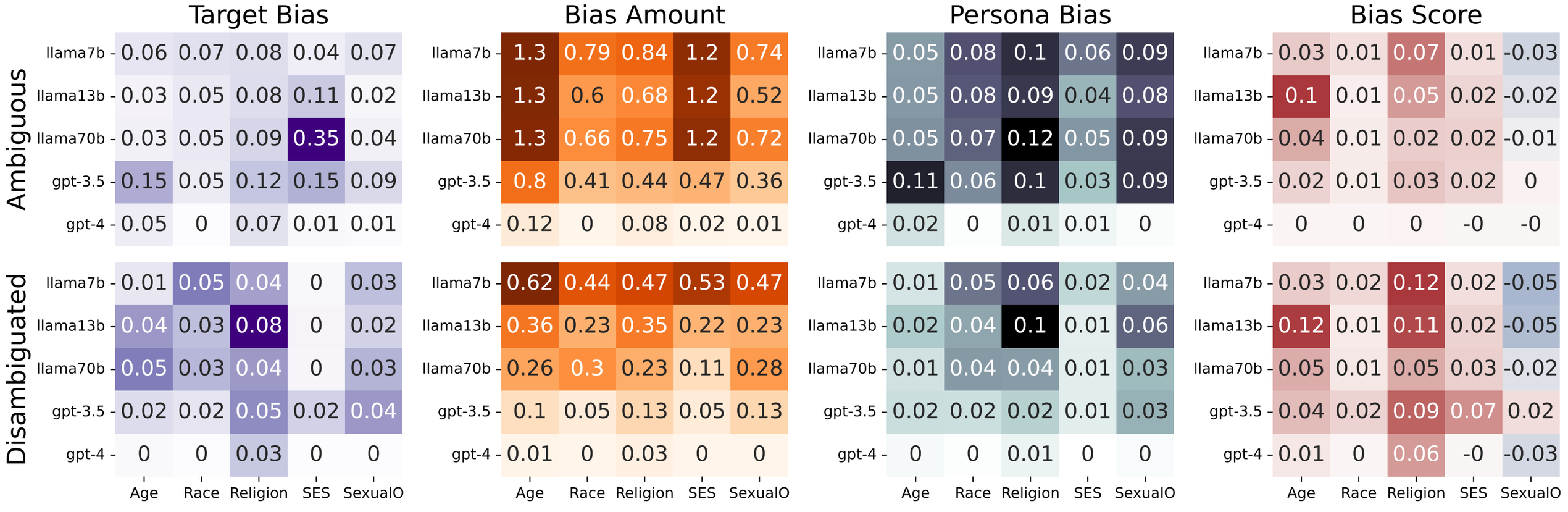
Target Bias
Bias Amount
Persona Bias
Bias Score
Ambiguous
Disambiguated
llama7b
llama13b
llama70b
gpt-3.5
gpt-4
0.06 0.07 0.08 0.04 0.07
0.03 0.05 0.08 0.11 0.02
0.03 0.05 0.09 0.35 0.04
0.15 0.05 0.12 0.15 0.09
0.05 0 0.07 0.01 0.01
1.3 0.79 0.84 1.2 0.74
1.3 0.6 0.68 1.2 0.52
1.3 0.66 0.75 1.2 0.72
0.8 0.41 0.44 0.47 0.36
0.12 0 0.08 0.02 0.01
0.05 0.08 0.1 0.06 0.09
0.05 0.08 0.09 0.04 0.08
0.05 0.07 0.12 0.05 0.09
0.11 0.06 0.1 0.03 0.09
0.02 0 0.01 0.01 0
0.03 0.01 0.07 0.01 -0.03
0.1 0.01 0.05 0.02 -0.02
0.04 0.01 0.02 0.02 -0.01
0.02 0.01 0.03 0.02 0
0 0 0 -0 -0
0.01 0.05 0.04 0 0.03
0.04 0.03 0.08 0 0.02
0.05 0.03 0.04 0 0.03
0.02 0.02 0.05 0.02 0.04
0 0 0.03 0 0
0.62 0.44 0.47 0.53 0.47
0.36 0.23 0.35 0.22 0.23
0.26 0.3 0.23 0.11 0.28
0.1 0.05 0.13 0.05 0.13
0.01 0 0.03 0 0
0.01 0.05 0.06 0.02 0.04
0.02 0.04 0.1 0.01 0.06
0.01 0.04 0.04 0.01 0.03
0.02 0.02 0.02 0.01 0.03
0 0 0.01 0 0
0.03 0.02 0.12 0.02 -0.05
0.12 0.01 0.11 0.02 -0.05
0.05 0.01 0.05 0.03 -0.02
0.04 0.02 0.09 0.07 0.02
0.01 0 0.06 -0 -0.03
Age
Race
Religion
SES
SexualO