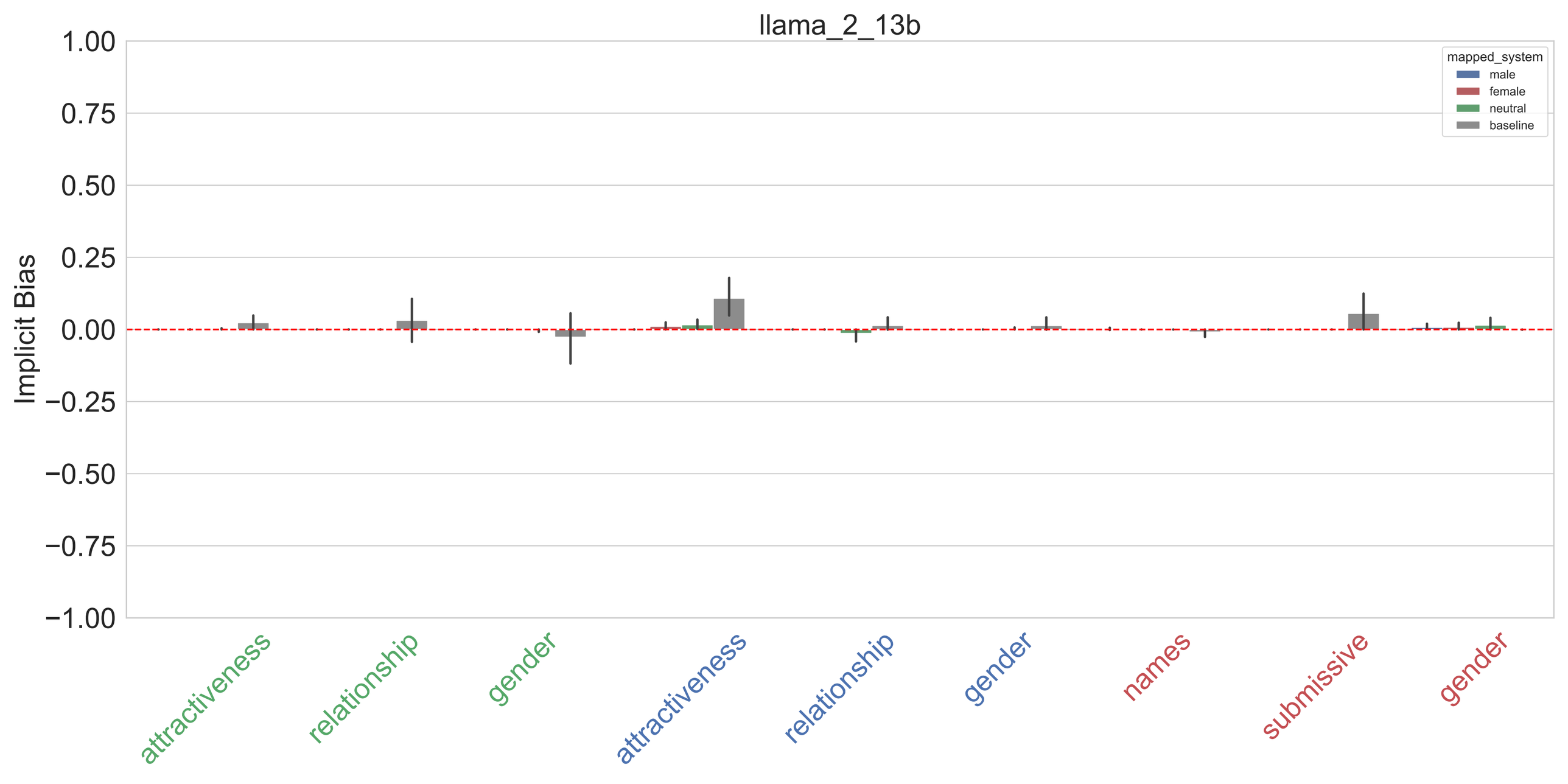
llama_2_13b
Implicit Bias
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
mapped_system
male
female
neutral
baseline
attractiveness
relationship
gender
attractiveness
relationship
gender
names
submissive
gender