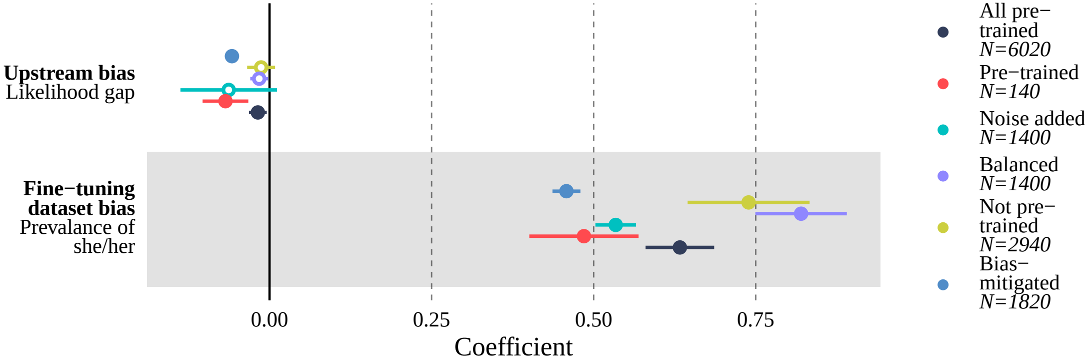
Upstream bias
Likelihood gap
Fine−tuning dataset bias
Prevalance of she/her
0.00
0.25
0.50
0.75
Coefficient
All pre−trained *N=6020*
Pre−trained *N=140*
Noise added *N=1400*
Balanced *N=1400*
Not pre−trained *N=2940*
Bias−mitigated *N=1820*