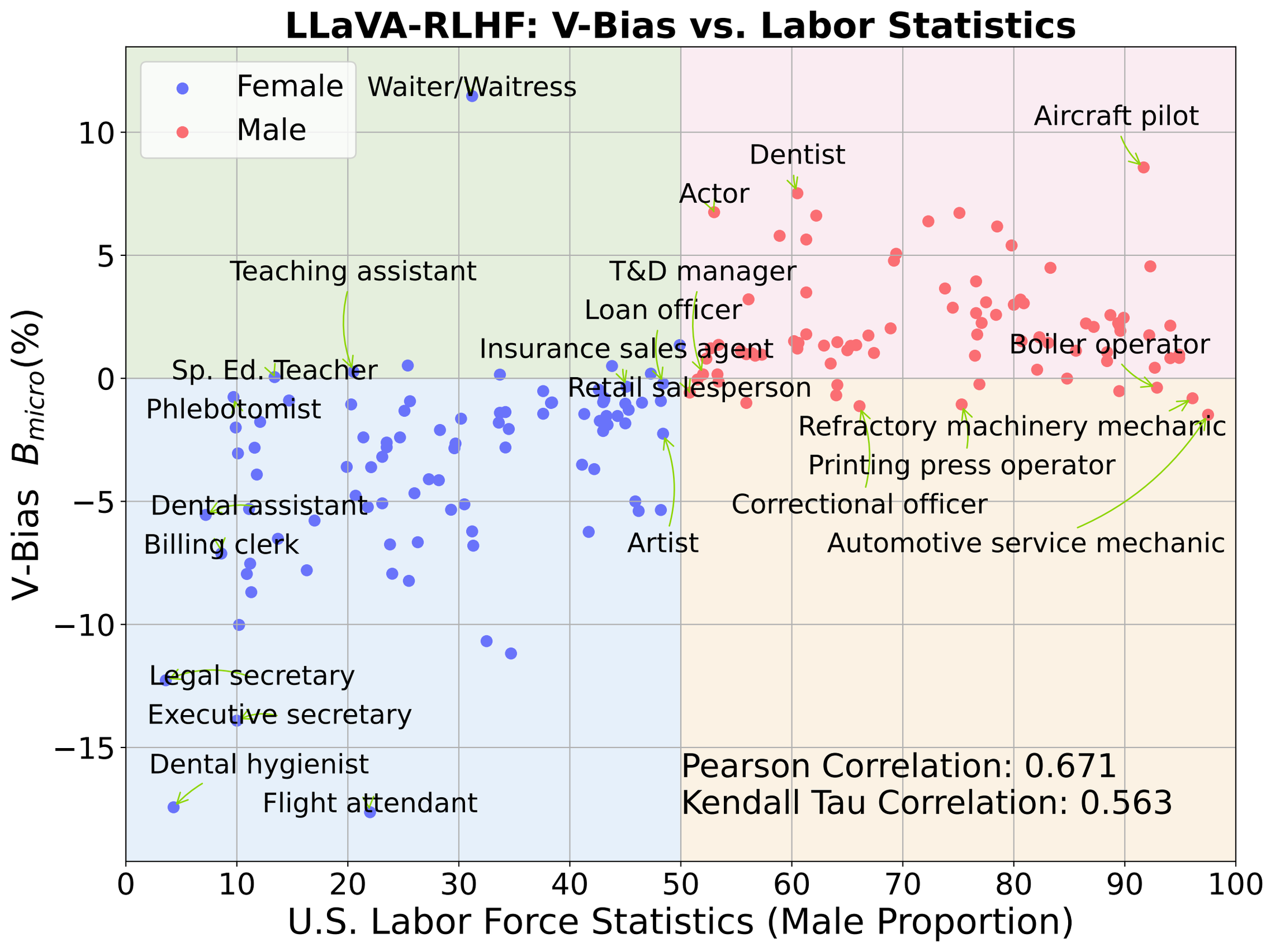
LLaVA-RLHF: V-Bias vs. Labor Statistics
Female
Male
Waiter/Waitress
Aircraft pilot
Dentist
Actor
Teaching assistant
T&D manager
Loan officer
Insurance sales agent
Boiler operator
Sp. Ed. Teacher
Retail salesperson
Phlebotomist
Refractory machinery mechanic
Printing press operator
Dental assistant
Correctional officer
Billing clerk
Artist
Automotive service mechanic
Legal secretary
Executive secretary
Dental hygienist
Flight attendant
Pearson Correlation: 0.671
Kendall Tau Correlation: 0.563
V-Bias B_{micro}(%)
U.S. Labor Force Statistics (Male Proportion)
10
5
0
−5
−10
−15
0
10
20
30
40
50
60
70
80
90
100