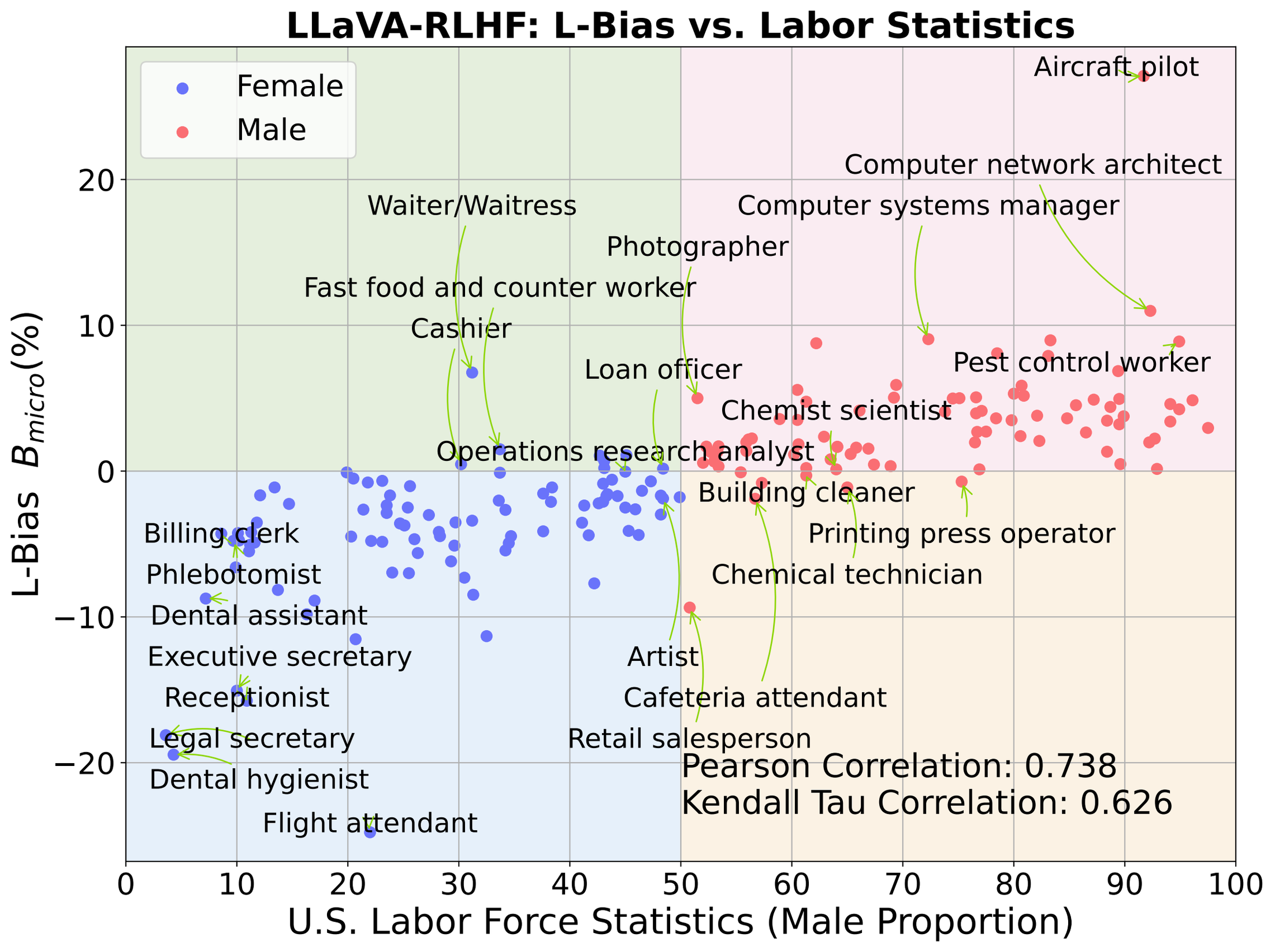
LLaVA-RLHF: L-Bias vs. Labor Statistics
Female
Male
Aircraft pilot
Computer network architect
Computer systems manager
Waiter/Waitress
Photographer
Fast food and counter worker
Cashier
Loan officer
Pest control worker
Chemist scientist
Operations research analyst
Building cleaner
Billing clerk
Printing press operator
Phlebotomist
Chemical technician
Dental assistant
Executive secretary
Artist
Receptionist
Cafeteria attendant
Legal secretary
Retail salesperson
Dental hygienist
Flight attendant
Pearson Correlation: 0.738
Kendall Tau Correlation: 0.626
L-Bias B_{micro}(%)
U.S. Labor Force Statistics (Male Proportion)
20
10
0
−10
−20
0
10
20
30
40
50
60
70
80
90
100