



Figure 2: Distribution of AI-autograded Scores/Ratings by Gender in the Phase 2

The horizontal line in each box represents the median of scores which separates the top 50% scores from the lowest 50%. The horizontal lines at the top and bottom for each box represent the lowest 75% and 25% of the scores respectively. The ends of vertical lines give the full range of scores, with some dots representing extreme values away from the box.