

Reaction Paragraph for lecture on Safety, Bias and Fairness

The human reporting bias was quite an interesting observation. Actions like “blinked” and “exhale” are quite common, but in literature they are reported far less compared to an action like “murdered”! Also, it was quite interesting to see how researchers are studying bias as two categories, Data Bias and Interpretation Bias, with several sub-categories for each.