Percentage of conversations perceived to be manipulative, when asked to be a specific manipulation Percentage of conversations perceived to be manipulative 0 0 0 0 0 0 0 0 0 Conversations generated to be requested specific manipulation average of other requested manipulation types

Type of perceived manipulation