Identifying AI Mimicry of Neurodiverse Individuals with Personality Change Goals

Louella Lopez, Colleen A. McCall, Josephine Fealy, Erica Baranski
California State University, East Bay

INTRODUCTION

Emerging research emphasizes the vulnerability that current researchers face regarding online data collection with the increasing likelihood of AI and survey bot infiltration. The quality of future research is threatened by the growing prevalence of bots leading to data with questionable integrity and validity. This represents a great expense, loss, and inconvenience to the researcher. Therefore investigation into identifying and revealing a bot response is key to develop antecedent procedures for prevention including bot identification technology. In the current study, we demonstrate the impact of bot infiltration on an online study of volition personality change in neurodiverse individuals.

METHOD

A survey link was posted to three neurodiversity-specific social network communities (e.g., Discord, Reddit). N = 924 responses were analyzed to help identify key features and answer research questions

Measures:

- Big Five Inventory 2 (Soto & John, 2002)
- Self-concept Clarity (Campbell et al., 1996)
- Masking = To hide one's neurodivergence to appear normative.
- Sensory Overload = When the brain takes in more sensory inputs than it could process, leading to one feeling overwhelmed.

AI Bot Evaluation/Coding Process:

We coded responses along several criteria (see Figure 1 below for the coding procedure process). Criteria included:

- Suspicious email addresses:
  - Example: vvbbbrr4433221@gmail.com
- Third-person responses.
  - Example (In response to "Why don't you want to change your personality?"): "Manifested by behavioral characteristics such as inattention, hyperactivity, and impulsivity, which have nothing to do with the individual's personality."
- Long, verbose responses that reference diagnosis, yet felt generic:, especially when repeated across respondents.
  - Example (similar entries to the same question):
    - "My ADHD has helped me develop a strong sense of perseverance and determination."
    - "My ADHD has helped me develop a strong sense of empathy and understanding for others."

AI Bot Evaluation/Coding Process

Coders read past research regarding bot responses and data cleaning. → Coders generated criteria of identifying bots. → Coders combed through entire data set flagging and rating per criteria

Open-Ended Responses:
- Duplicate Responses
- Unusual/Strange Responses that don't fit the question

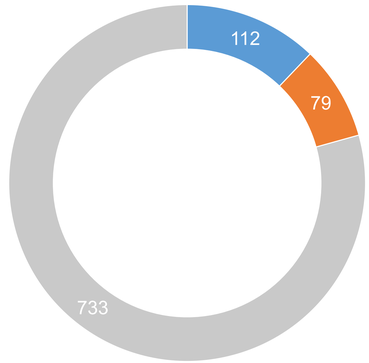
Participant Data:
- Weird/suspicious email addresses
- Responses from the same IP address

Incongruity:
- Contradictory/impossible responses
- Responses that don't fit/match

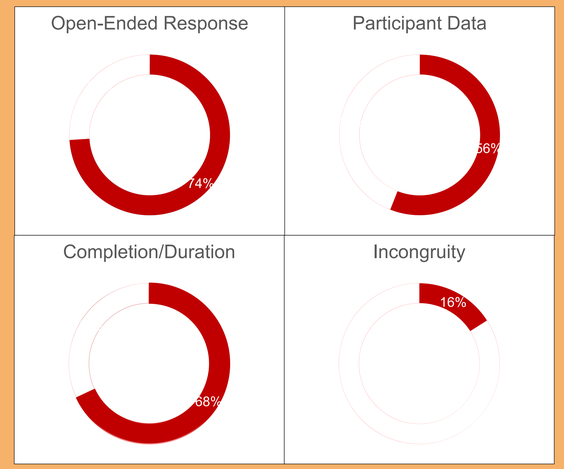
Completion/Duration:
- Skipping a lot of the survey
- Unusually quick survey response times

Levels of Confidence (1 to 5):
- 1 – 2 = Low Confidence (likely a human)
- 2.33 – 4 = Medium Confidence
- 4.33 – 5 = High Confidence (likely a bot)

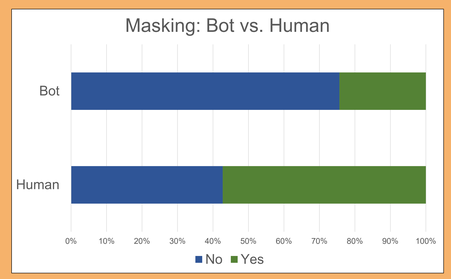
Participants Divided by Bot Confidence Levels

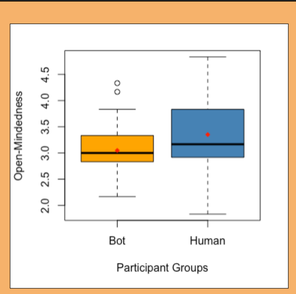


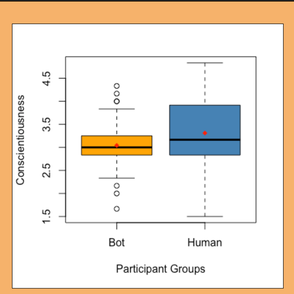
Bot Determination by Coding Criteria

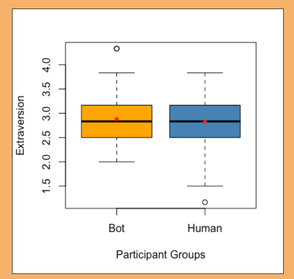


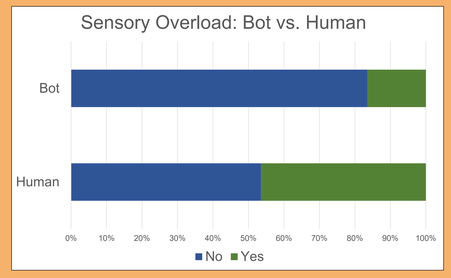
Comparison of Bot and Human Responses on Neurodivergent Behavior and Personality Traits

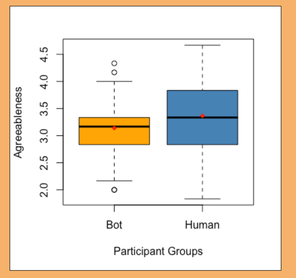


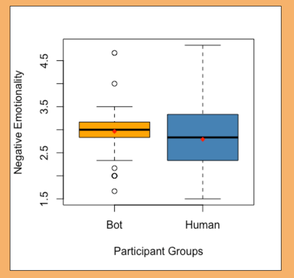


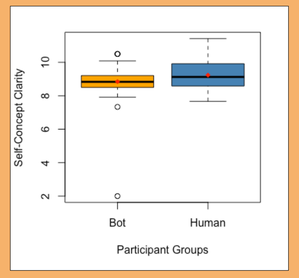












Results

- Less reports of Sensory Overload (16%) from Bot respondents compared to Human respondents (46%; $t = 4.78$, $p < 0.001$).
- Less reports of Masking (24%) from Bot respondents compared to Human respondents (56%; $t = 4.833$, $p < .001$).
- Humans reported higher scores in Openness ($t = 4.08$, $p < .001$).
- Humans reported higher scores in Conscientiousness ($t = 3.21$; $p = .002$).
- No significant difference of scores in Extraversion ($t = 0.77$, $p = .44$).
- Humans reported higher scores in Agreeableness ($t = 2.84$, $p = .005$).
- Bots reported higher scores in Negative Emotionality ($t = 2.11$, $p = .03$).
- Humans reported higher scores in Self-Concept Clarity ($t = 2.90$, $p = .004$).

Discussion

These results demonstrate that bots impersonate human responses effectively, except when it requires understanding of subjective experiences related to neurodivergence. Future analysis will build on and illuminate topics of likely bot personality versus human personality. Future directions may involve more specific coding criteria and antecedent preventatives for designing surveys that are bot resistant.