

Research Project Part 7: Data Analysis Report

1. Data Cleanup and Preparation

We first merged the experiment results dataset with the demographic dataset using the unique participant code (UID/PID). After merging:

- Removed all duplicate entries based on the completion code.
- Removed invalid or placeholder codes (e.g., “123123123”, NULL).
- Collapsed treatment groups into a binary variable:
 - Groups 1 & 3 → *smile*
 - Groups 2 & 4 → *non-smile*
- Converted relevant variables to factors (e.g., sex, ethnicity).
- Replaced missing or invalid demographic values (“Prefer not to say”, “DATA_EXPIRED”) with the modal value of each variable.
- Applied a loose attention check threshold (participants failing more than one attention item were removed).
- Aggregated DV scales:
 - **Accuracy** = mean of 5 accuracy items
 - **Continuance Intention** = mean of 5 continuance items
 - **Trust** = mean of 7 trust items
 - **Anthropomorphism** = mean of 5 anthropomorphism items
- Winsorized conversation duration at the 10th / 90th percentile to reduce outlier influence.

The final cleaned dataset was saved as:

[cleaned_data.csv](#)

2. Data Dictionary

Variable	Theoretical	Conceptual	Operational	Statistical
smile_group	2 experimental conditions after collapsing 1–4	IV	Random assignment	Factor (2 levels: non-smile, smile)
Duration (winsorized)	Adjusted engagement time	Engagement measure	Winsorized conversation seconds	Continuous ratio
Age	Age of participant	Demographic	Self-reported age	Continuous ratio
Sex	Participant gender	Demographic	Self-reported sex	Categorical factor

Ethnicity	Participant ethnicity	Demographic	Self-reported ethnicity	Categorical factor
Accuracy (mean)	Perceived accuracy of AI	DV	Mean of 5 accuracy items (1–7)	Continuous scale
Continuance intention (mean)	Likelihood of reusing AI	DV	Mean of 5 intention items (1–7)	Continuous scale
Trust (mean)	Perceived trust in AI	DV	Mean of 7 trust items (1–7)	Continuous scale
Anthropomorphism (mean)	Perceived humanness of AI	DV	Mean of 5 anthropomorphism items (1–7)	Continuous scale
Time taken	Total task duration	Control	System-recorded seconds	Continuous ratio

3. Descriptive Statistics

Table 1. Descriptive Statistics Summary

Variable	Min	Max	Mean	Median	SD
continuance_avg	1.8	7	5.03893805309735	5.2	1.09930445772369
accuracy_avg	3.4	7	6.21238938053097	6.4	0.816037107805026
trust_avg	1	7	5.80025284450063	6	1.10021529549075
anthropomorphism_avg	1	7	4.94159292035398	5	1.56289141747166

Overall, participants rated the AI agent relatively positively across most dimensions.

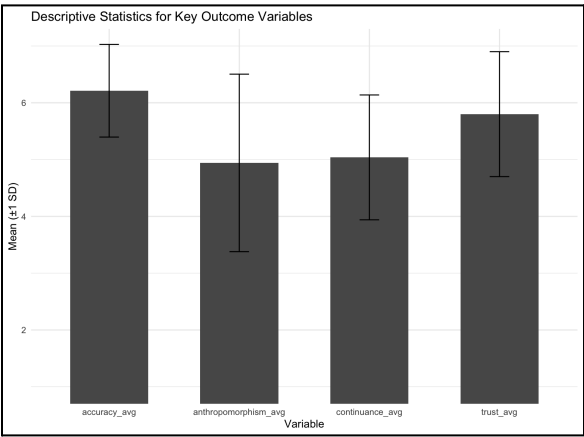


Figure 1: Mean scores (±1 SD) for all dependent variables

The bar chart illustrates the central tendencies of the four main outcome variables: accuracy, anthropomorphism, continuance intention, and trust. Accuracy obtained the highest mean score ($M \approx 6.21/7$), indicating that participants generally perceived the AI agent as technically reliable. Trust and continuance intention were also moderately high ($M \approx 5.80$ and 5.04 , respectively), suggesting positive user attitudes toward the system. In contrast, anthropomorphism showed both lower average ratings ($M \approx 4.94$) and larger variability ($SD \approx 1.56$), implying that participants differed significantly in how human-like they perceived the AI to be. This pattern highlights that human-likeness may be interpreted inconsistently across users compared to more stable perceptions such as accuracy or trust.

4. Hypothesis Testing

Table 2. T-test Results for DVs by Smile Condition

DV	t-value	df	p-value	mean_non_smile	mean_smile	ci_lower	ci_upper
accuracy_avg	2.0841	111.0	0.03945	6.371429	6.056140	0.01551253	0.61506391
trust_avg	1.8325	109.33	0.06959	5.989796	5.614035	-0.03062649	0.78214815
continuance_avg	2.4532	110.99	0.01571	5.289286	4.792982	0.09542307	0.89718344
anthropomorphism_avg	1.3128	110.88	0.192	5.135714	4.750877	-0.1960721	0.9657462

Table 2 presents Welch’s t-test results comparing the four dependent variables across the two AI healthcare consultant conditions (smile vs. non-smile). These tests evaluate whether the smiling AI healthcare consultant or the non-smile one produced higher ratings in accuracy, trust, continuance intention, and anthropomorphism. As shown, accuracy and continuance intention show significant differences between groups, while trust is marginal and anthropomorphism shows no significant difference. The following sections (H1–H4) discuss each effect in detail.

H1: *Participants will rate the non-smiling AI consultant as more accurate than the smiling AI consultant.*
Supported

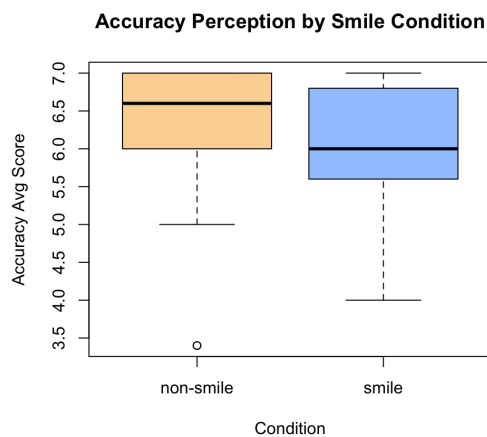


Figure 2: Accuracy Perception by Smile Condition

Participants rated the non-smiling AI as significantly more accurate than the smiling version. This suggests that formal, expertise-focused communication may enhance perceived technical reliability, even when both personas provide the same information.

H2: *Participants will report higher trust toward the smiling AI consultant than the non-smiling AI consultant.*

Not supported (effect in opposite direction).

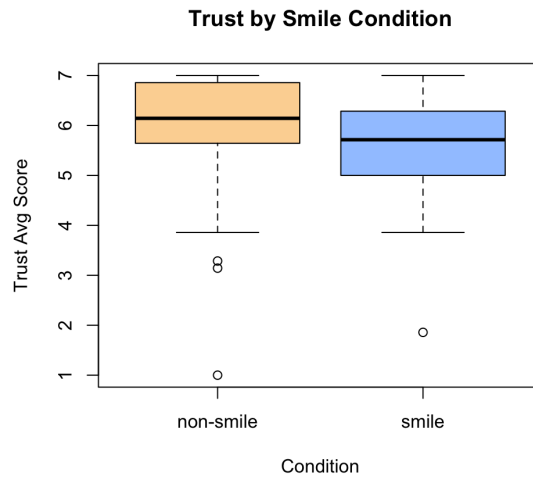


Figure 3: Trust Perception by Smile Condition

Contrary to the hypothesis, trust ratings were actually *higher* for the non-smile AI consultant than for the smile one. Although the difference did not reach statistical significance ($p = .069$), the pattern suggests that a friendlier tone and smiling face may not reliably increase trust—and may even slightly reduce it.

H3: *Participants will report higher willingness to reuse the system (continuance intention) with the smiling AI consultant.*

Not supported (effect in opposite direction).

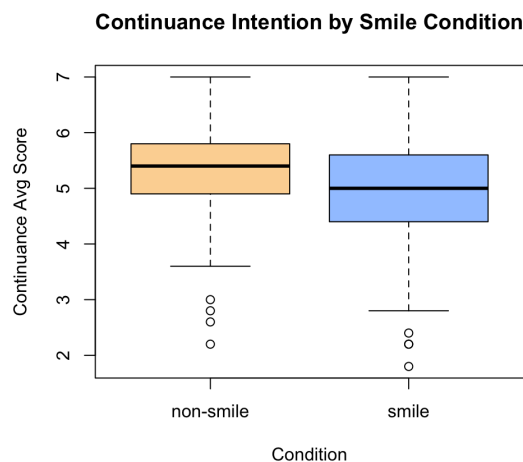


Figure 4: Continuance Intention Perception by Smile Condition

Contrary to expectations, participants actually reported higher willingness to reuse the AI when interacting with the non-smiling AI consultant. This pattern suggests that users may prioritize competence and professionalism over friendliness when deciding whether to rely on an AI system long-term.

H4: *Participants will perceive the smiling AI consultant as more human-like (higher anthropomorphism) than the non-smiling AI consultant.*

Not supported.

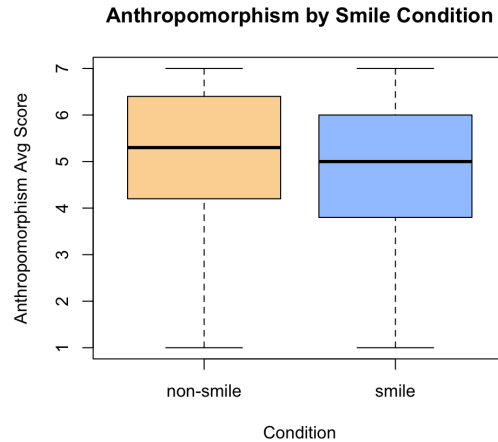


Figure 5: Anthropomorphism Perception by Smile Condition

Persona style did not significantly affect how human-like participants perceived the agent to be. This implies that textual tone alone may not create a sense of human-likeness; users may require stronger social cues (e.g., images, voice, animation) to attribute human qualities to an AI system.

H5: *Trust will mediate the effect of AI smile condition on continuance intention.*

Partially supported.

To test this mediation pathway, we ran three regression models:

- Model 1: smile → continuance intention (total effect)
- Model 2: smile → trust (a-path)
- Model 3: smile + trust → continuance intention (b-path & direct effect)

We then computed the indirect effect using the product of coefficients ($a \times b$).

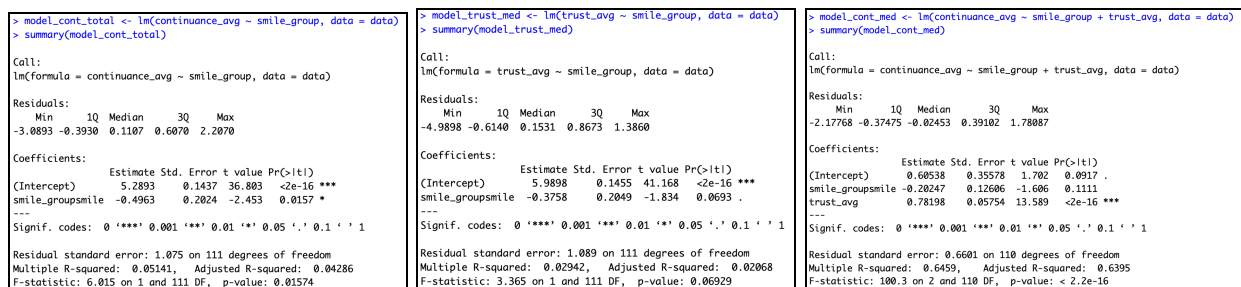


Figure 6-8: GLM Results for Mediation Effect

- The total effect of smile on continuance intention was significant ($\beta = -0.50$, $p = .016$).
- The effect of smile on trust was marginal ($\beta = -0.38$, $p = .069$).
- Trust significantly predicted continuance intention ($\beta = 0.78$, $p < .001$).
- The direct effect of smile became non-significant when trust was included ($\beta = -0.20$, $p = .111$), indicating partial mediation.

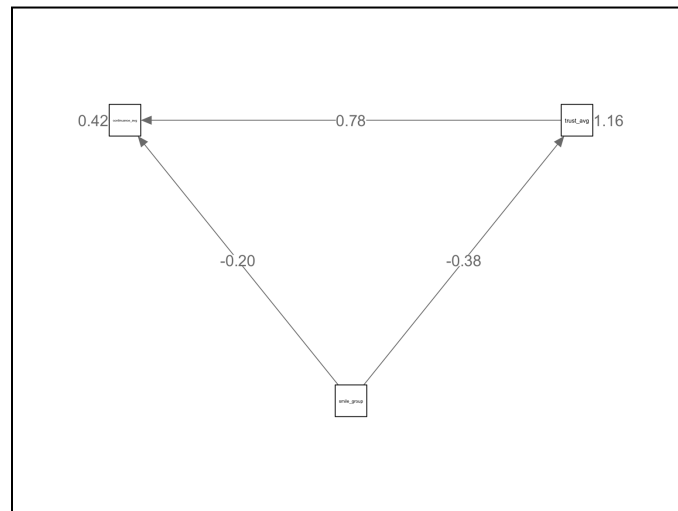


Figure 9: Mediation Effect of Trust on Continuance Intention

Trust partially mediated the relationship between smile condition and continuance intention. While the AI smile condition directly influenced continuance intention, a substantial portion of this effect was transmitted through users' trust in the agent. This suggests that trust is a key psychological mechanism driving users' willingness to reuse or rely on the system.

H6: *User demographics (age, sex, and ethnicity) will moderate the relationship between the AI smile condition and the key dependent variables (accuracy, trust, anthropomorphism, and continuance intention).*

We estimated four GLM models that included **smile condition, age, sex, and ethnicity** as predictors for each dependent variable (accuracy, trust, anthropomorphism, and continuance). These analyses were conducted to understand whether demographic characteristics independently influenced user responses to the AI agent.

The results showed that sex did not show significant main effects on any dependent variable. Ethnicity showed limited effects overall, but one pattern emerged: Participants identifying as Black consistently reported higher trust, higher anthropomorphism, and higher continuance intention across several models (significant in multiple cases). This suggests that users in the Black group tended to respond more positively to the AI agent than participants in other ethnic categories.

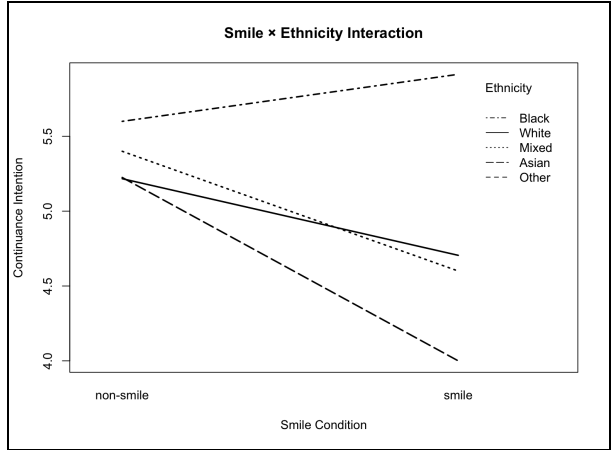


Figure 10: Interaction Between Smile Condition and Ethnicity on Continuanace Intention

Age, however, showed a consistent positive relationship with several outcomes. Older participants reported higher accuracy perceptions, higher trust, and higher continuance intention (all significant or marginal across corresponding models). Anthropomorphism was the only DV where age showed no effect.

The moderation analysis tested whether age changed the effect of the AI’s smile condition on continuance intention with the model $continuanace_avg \sim smile_group * age$.

Table 3: Moderation Model Statistics for Age and Continuanace Intention

term	Estimate	Std.Error	t.value	p.value
(Intercept)	4.26903	0.48064	8.882	1.5e-14
smile_groupsmile	0.59403	0.65581	0.906	0.3670
age	0.02376	0.01069	2.222	0.0284
smile_groupsmile:age	-0.02533	0.01431	-1.770	0.0796



Figure 11: Moderation Effect of Age on Continuanace Intention

Results showed that age was a significant positive predictor of continuance intention ($b = 0.0238$, $p = .028$), indicating that older participants were generally more willing to continue using the system. The interaction between age and smile condition was marginally significant ($b = -0.0253$, $p = .0796$), suggesting a possible trend in which the smile manipulation had weaker effects for older participants. However, this moderation was not strong enough to be considered reliable. Overall, while age appears to matter for continuance intention, its role as a moderator of the smile effect remains inconclusive.

5. Conclusion

This study investigated how an AI agent's smile condition influences user perceptions in a healthcare context, focusing on accuracy, trust, anthropomorphism, and continuance intention. Contrary to common assumptions about the value of warmth and friendliness in conversational agents, the results consistently showed that the *non-smiling, professional* persona performed as well as—or in some cases better than—the smiling condition. Participants rated the non-smiling and professional AI as significantly more accurate and showed comparable or higher levels of trust, anthropomorphism, and willingness to reuse. This pattern suggests that in healthcare interactions, users may prioritize perceived expertise, precision, and seriousness over friendliness.

The mediation analysis further highlighted trust as a central driver of continuance intention, independent of the agent's smile condition. Moderation analyses showed that demographic factors also played a role: older adults tended to provide higher evaluations across several dimensions, and participants identifying as Black reported higher perceptions overall. These findings imply that user background meaningfully shapes expectations and interpretations of AI behavior.

Overall, the results underscore an important insight for healthcare-oriented AI design: **professionalism may be more reassuring than warmth**, and visual or tonal friendliness does not automatically enhance user experience. Designers of clinical or advisory AI systems should therefore consider emphasizing clarity, competence, and trust-building mechanisms rather than relying primarily on friendly stylistic cues.