

Talking Like a Human: How Conversational Anthropomorphism Affects Self-Disclosure to Mental Health Chatbots

Author: Sagar Chethan Kumar Supervisor: Esra de Groot Responsible Professor: Prof. Dr. Ujwal Gadiraju

Delft University of Technology

Background & Related Work

- AI chatbots offer accessible and non-judgmental spaces for support [1, 7].
- A core challenge is fostering *self-disclosure*, critical for tailored help but shaped by how chatbots communicate [16].
- Conversational anthropomorphism**—e.g., humor, small talk—can boost disclosure in casual settings [2, 19], but may backfire in sensitive contexts [15, 20].
- Effects depend on **question sensitivity**, or how intrusive users perceive a question to be [16].
- This interaction is underexplored in mental health chatbot design [12, 11].

Research Questions & Hypotheses

How do conversational anthropomorphism and question sensitivity influence self-disclosure to AI-powered mental health chatbots?

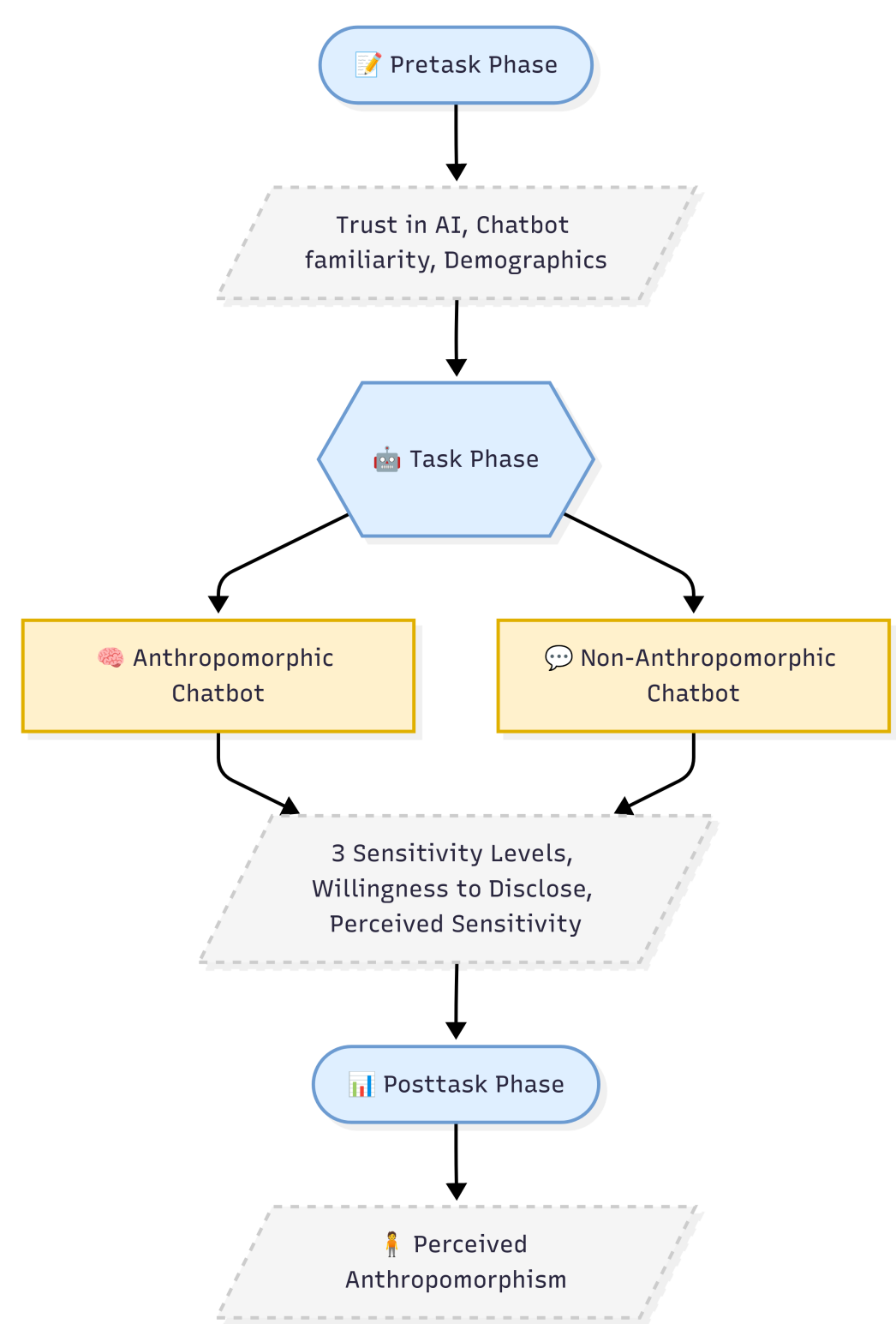
Research Questions

- RQ1.** Does conversational anthropomorphism increase willingness to self-disclose?
- RQ2.** Does the sensitivity of disclosure-intent questions influence willingness to self-disclose?
- RQ3.** Is there an interaction between conversational anthropomorphism and question sensitivity on willingness to self-disclose?

Hypotheses

- H1.** Conversational anthropomorphism will increase users' willingness to self-disclose [12, 11].
- H2.** Users will be less willing to self-disclose as the sensitivity of questions increases [15, 2].
- H3.** Conversational anthropomorphism will increase willingness to self-disclose for low-sensitivity questions but decrease willingness for high-sensitivity questions [16, 11, 2, 15].

Experimental Setup



Variables:

- Independent:** Anthropomorphism (low vs. high), Question Sensitivity (low, medium, high)

- Dependent:** Willingness to Self-Disclose

Measures:

- Pre-Task:** Trust in AI [8], Chatbot Familiarity [5], Age (5 year bins), and Gender
- During Task:** Self-Disclosure Willingness, Perceived Sensitivity
- Post-Task:** Perceived Anthropomorphism (manipulation check) [3, 10]

Participants (n=30):

- Gender: 60% male, 40% female
- Age: 21–25 (53.3%), 16–20 (33.3%), 26–30 (13.3%)

Methodology

Operationalization: Few-shot prompting [18] for persona adoption; embedded conversational styles [9].

- Control (Non-Anthropomorphic):** Neutral, formal tone; factual, non-adaptive responses [20, 15].
- Experimental (Anthropomorphic):** Informal tone, emojis, humor, typing delays and indicators; adapted to user willingness and topic sensitivity [2, 16].

Questions: Three randomized scenarios with general dialogue and three fixed disclosure questions (low, medium, high sensitivity) from the **SelfDisclosureItems** dataset [13]. Participants reported willingness rather than actual disclosures.

Analysis Plan:

- Factorial mixed ANOVA:** Anthropomorphism (between-subjects) \times Sensitivity (within-subjects); 4 confound controls [14].
- Manipulation Check:** Independent-samples *t*-test on perceived anthropomorphism.
- Effect Sizes and Confidence Intervals:** ω^2 , η^2 , η_G^2 , Hedges' *g* [6, 4].
- Assumption Checks:**
 - Normality: Shapiro–Wilk, Q–Q plots
 - Sphericity: Mauchly's test (G–G or H–F correction)
 - Variance homogeneity: Levene's test (Welch's ANOVA if violated) [17]

Results

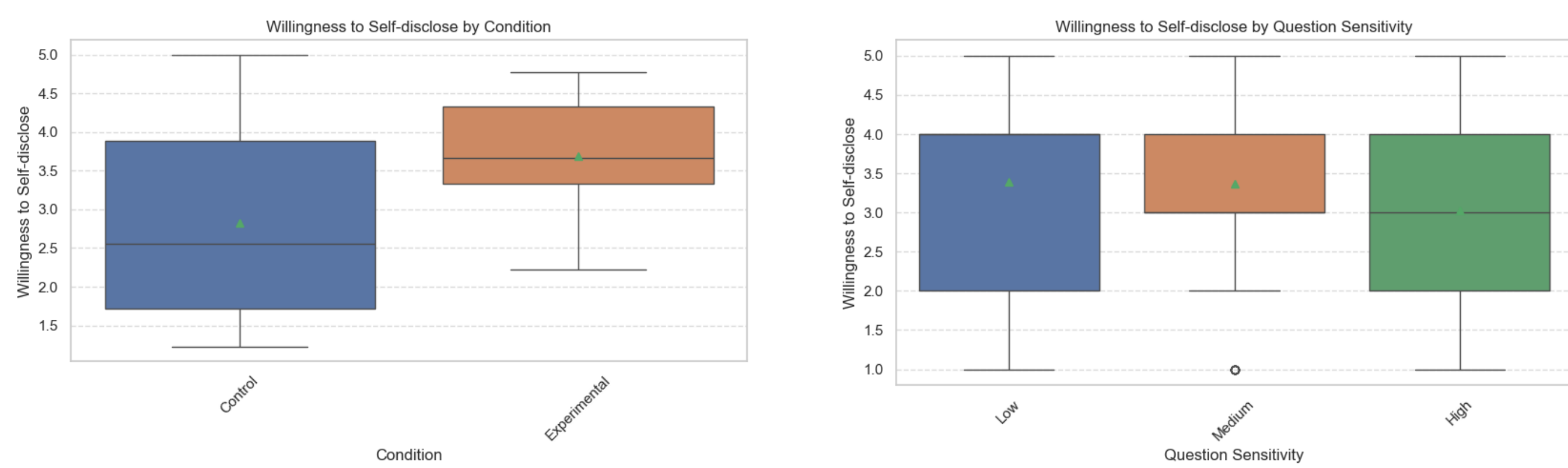


Figure 1. Willingness to self-disclose by condition (left) and by sensitivity (right).

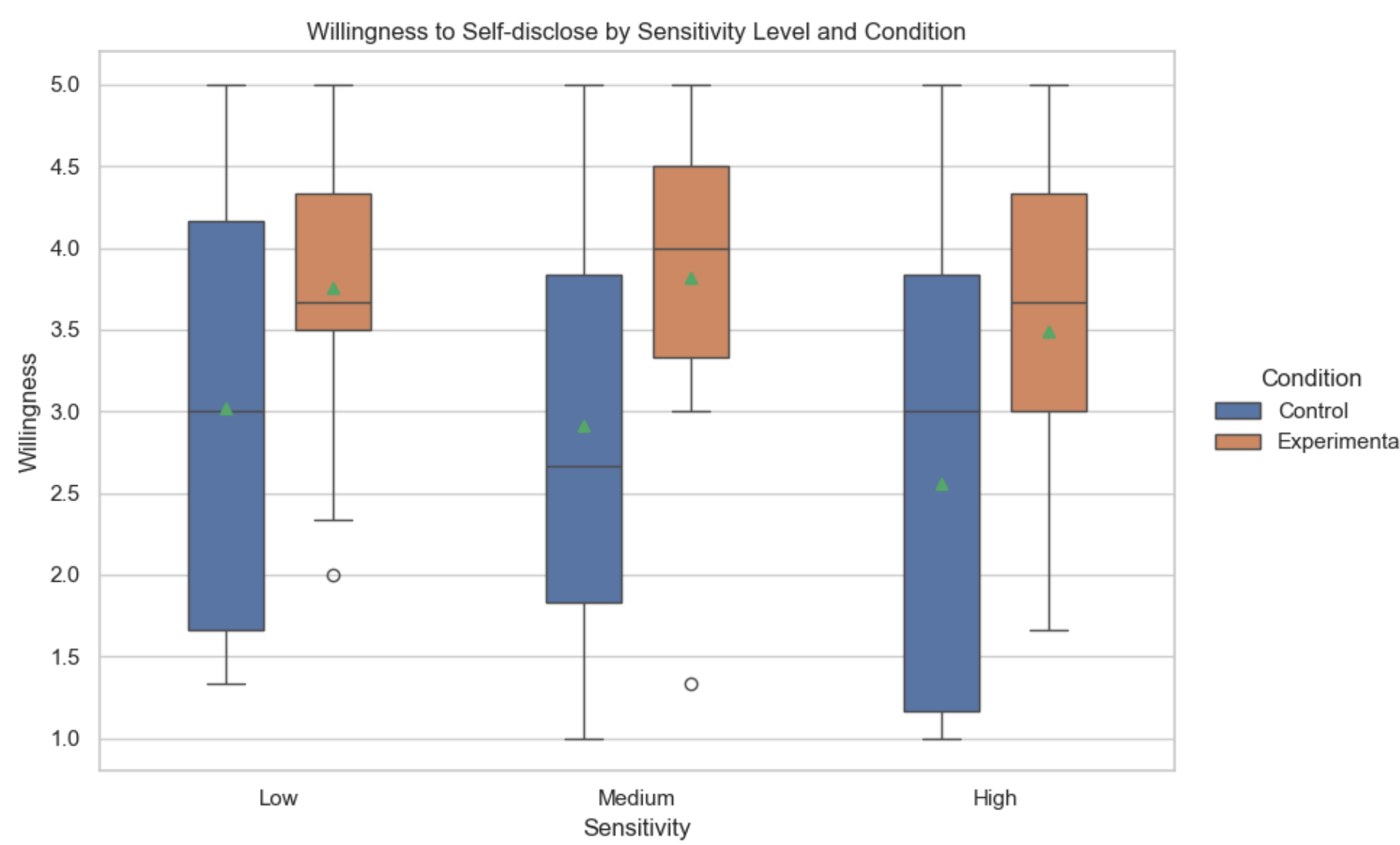


Figure 2. Willingness to self-disclose across condition and question sensitivity.

Discussion

RQ1. Descriptive trends show higher willingness to self-disclose with anthropomorphic chatbots across all sensitivity levels, partially supporting **H1**. However, this effect was *not statistically significant*, suggesting anthropomorphism alone may not meaningfully shift behavior in sensitive contexts.

RQ2. Willingness decreased as question sensitivity increased, aligning with **H2**, but this trend also lacked statistical significance after correction, possibly due to limited power.

RQ3. No significant interaction was found. However, under anthropomorphic conditions, willingness remained higher across sensitivity levels, tentatively supporting **H3** and potentially suggesting reduced evaluative concerns of topic sensitivity in mental health.

Limitations

- Small Sample Size ($n = 30$):** Limited power and increased Type II error risk.
- Self-Reported Willingness:** May not reflect real-world disclosure behavior.
- Survey-Based Design:** Reduces ecological validity; lacks depth of open-ended or real-time interaction.
- Assumption Violations:** Minor violations of normality and variance homogeneity; non-parametric methods may be more robust.

Conclusion & Future Work

Conversational anthropomorphism showed consistent, but non-significant, positive effects on users' willingness to self-disclose to mental health chatbots across all question sensitivities. Question sensitivity was inversely related to disclosure, but also non-significant. No interaction effects emerged, though exploratory trends suggest anthropomorphic cues may support disclosure even with sensitive topics.

Future Work

- Use larger, more diverse samples for generalizability.
- Include open-ended interactions and behavioral measures.
- Apply mixed-methods designs to capture emotional nuance and trust formation.

References

Note: References are truncated for visual presentation. Full citations available upon request.

- Alaa A Abd-Alrazaq, Mohannad Alajlani, Ali Abdallah Alalwan, Bridgette M Bewick, Peter Gardner, and Mowafa Househ. An overview of the features of chatbots in mental health: A scoping review. *International journal of medical informatics*, 132:103978, 2019.
- Martin Adam and Johannes Klumpe. Onboarding with a chat—the effects of message interactivity and platform self-disclosure on user disclosure propensity. *European Conference on Information Systems*, 2019.
- Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1:71–81, 2009.
- Marc Brysbaert and Michaël Stevens. Power analysis and effect size in mixed effects models: A tutorial. *Journal of cognition*, 1(1):9, 2018.
- Astrid Carolus, Martin J Koch, Samantha Straka, Marc Erich Latoschik, and Carolin Wienrich. Mails-meta ai literacy scale: Development and testing of an ai literacy questionnaire based on well-founded competency models and psychological change-and meta-competencies. *Computers in Human Behavior: Artificial Humans*, 1(2):100014, 2023.
- Adam Claridge-Chang and Pryseley N Assam. Estimation statistics should replace significance testing. *Nature methods*, 13(2):108–109, 2016.
- MDR Haque and S Rubya. An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews. *jmir mhealth and uhealth*, 11, e44838, 2023.
- Jiun-Yin Jian, Ann Bisantz, and Colin Drury. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4:53–71, 03 2000.
- Ivica Kostrić, Krisztian Balog, and Ujwal Gadiraju. Should we tailor the talk? understanding the impact of conversational styles on preference elicitation in conversational recommender systems, 2025.