Humans relate more with LLMs that mirror their psychology

Santiago Castiello1,2 \*, Riddhi Jain Pitliya1\*, Daniel R. Lametti3,4\* , Robin A. Murphy2\*

1. Yale School of Medicine, Yale University
2. Department of Experimental Psychology, University of Oxford
3. Department of Psychology, Acadia University
4. OneReach.ai

\*all authors contributed equally

Keywords: homophily, affiliation, large language models, AI, human-AI interaction

# Abstract

People connect with others who share their likes, interests, and mental states, a phenomenon known as homophily. Here, we examine human-AI homophily by testing how individuals' psychological traits shape their affiliation with large language models (LLM) that mimic their traits. In Experiment 1, the LLM LLM was instructed to produce language aligned with either an anxious or non-anxious state. Participants (n=100) engaged with both versions of LLM and then completed a questionnaire to assess how relatable they found the LLM. Participants with high anxiety felt more similar to and understood by the LLM instructed to mimic an anxious state, while participants with low anxiety felt different from it. In Experiment 2, participants (n=100) engaged with LLM instructed to produce language aligned with either an extroverted or introverted personality. Extroverted participants felt similar to the LLM that mimicked extroversion. Taken together, the findings support homophily in language-based human-AI interactions.

# 

# Introduction

Social interactions between individuals show effects of homophily (*homo*=similar*; philia*=love) and heterophily (*hetero*=different*; philia*=love) [(Lazarsfeld and Merton 1964)](https://paperpile.com/c/L1tWAD/Bi0B). These are the opposing tendencies to interact more easily and effectively with members of an affiliative group or with groups other than one’s own. Is homophily possible between AI and humans? The emergence of Large Language Models (LLMs) has expanded the range of groups with which we might affiliate to include artificial systems. As LLMs readily adopt specific linguistic styles in response to prompts, it is relatively easy to create person-centered or ‘tailored’ AI with the aim of building stronger human-AI bonds. But this idea—long a plot point for science fiction writers (see, for instance, the 2013 Spike Jonze film *Her*)—has little empirical support. Here, we prompted LLMs to use language associated with mental health related states (*i.e.*, anxious vs non anxious) or personality traits (*i.e.*, extraverted vs introverted) to systematically test whether human-AI homophily based on a shared psychology might occur. The experiments were designed to either enhance affiliative perceptions with AI among our participants or, conversely, negatively challenge them.

Turing’s thought experiment on the nature of human and artificial intelligence asked the question, “When must we acknowledge and grant artificial systems human-like intelligence? What must the artificial system do to be indistinguishable from a human conversational partner?” [(Turing 1950)](https://paperpile.com/c/L1tWAD/nVmX). Apart from a growing consensus that LLMs can now pass Turing’s test, or at least a basic version of this test, not surprisingly people have come to further refine definitions of intelligence, asking whether LLMs can reason or provide social comfort [(Mitchell 2024; Kambhampati 2024)](https://paperpile.com/c/L1tWAD/Hp0Y+zPIc). In spite of the engaging interactions LLMs offer, and their ability to answer complex knowledge and human reasoning questions, currently this system does not seem to pass the Turing test [(Jannai et al. 2023)](https://paperpile.com/c/L1tWAD/uLS8), however there are novel potential benchmarks to test General Intelligence that should be considered [(Biever 2023)](https://paperpile.com/c/L1tWAD/Yctb). We may notice that the AI misinterprets questions or intentions which may reflect true human interactions, but there are other factors that provide clues to the artificial nature of the experience. For example, they respond more quickly, too accurately, or not accurately enough [(Svenningsson and Faraon 2019)](https://paperpile.com/c/L1tWAD/7XAh). LLMs may also fail to meet our expectations, not because of lack of General Intelligence, but because we do not perceive an alignment with our personal characteristics. Specifically, LLMs may lack components that support an affiliative platform.

People exhibit homophily in their social networks—the tendency to affiliate with other individuals that share characteristics with themselves (Launay & Dunbar, 2015). Network relations are reflected in similar neural activity [(Parkinson et al. 2018)](https://paperpile.com/c/L1tWAD/cmYh). For example, people tend to choose partners who resemble themselves, a phenomenon observed in both in-person and app-based dating [(Restrepo-Echavarria et al. 2023)](https://paperpile.com/c/L1tWAD/vxkV) [(Furman and Simon 2008)](https://paperpile.com/c/L1tWAD/VEPg). Factors that contribute to homophily include things like shared religious beliefs, interests, hobbies, personality characteristics, mental states, and social network distance [(Parkinson et al 2018)](https://paperpile.com/c/L1tWAD/cmYh). But the degree to which people feel more similar to AI that share characteristics of their psychological traits remains unknown. Understanding the extent to which homophily is observed in human-AI interactions is important, not only for understanding how well human social behaviours generalize to intelligent computer systems, but also for the design of useful human-AI experiences (REFs). One practical application of large language might be in healthcare, specifically in the field of mental health and well-being. In this case, the degree to which humans affiliate with AI may dictate the usefulness of large language models as diagnostic or therapeutic tools (REFs).

The aim of the current study was to experimentally test whether humans feel more similar to an LLM that, through conversation, exhibits a shared psychological trait, despite full knowledge that they are talking to a computer program. In Experiment 1, the LLM GPT-4 was prompted (using clinically relevant criteria) to produce language aligned with either an anxious or non-anxious psychological state. One hundred participants conversed with both versions of GPT-4 and then answered a series of questions designed to test the extent to which they felt similar or different from the AI. Following this questionnaire, participants’ own anxiety was measured. We then examined participants’ sense of similarity to each GPT-4 persona in relation to their reported anxiety levels. As a further test of affiliation, we also examined how the sentiment of the messages they exchanged with each version of GPT-4 was influenced by the LLM’s language. We predicted that participants with high anxiety would feel more similar to the LLM that used anxious language and vice versa for participants with low anxiety, a key indicator of homophily.

In Experiment 2, we again examined the extent to which participants might feel similar to an LLM with a shared psychological trait, but we manipulated the personality trait of extroversion-introversion. GPT-4 was prompted (using answers to validated tests of personality) to produce language aligned with either an extroverted or introverted personality. One hundred participants had a conversation with each version of GPT-4; they then answered a series of questions designed to test the extent to which they felt similar or different from each AI. In this case, participants’ affiliation to each GPT-4 persona was examined in relation to reported extraversion. Again, we predicted homophily: extroverted participants would feel more similar to the GPT-4 persona that used extroverted language and vice versa for introverted participants.

Designing chatbots to provide more accessible information is important as a first goal, but ultimately people could use these platforms for social support, friendship but maybe therapy and diagnosis.

Dating apps present contundent evidence that via text messages people heuristics tend to be influenced by homophily. Alternatively people may be averse to interacting with those that differ from one’s position. Either tendency would encourage homophilic interactions. We hypothesised that participants with higher levels of anxiety (Experiment 1) or extraversion (Experiment 2) would engage more, and engage more effectively as indexed by self-disclosure, with AIs that the chatbot when the chatbot also exhibits similar anxious behaviours and beliefs. And that this might be accompanied by negative or weaker evaluations of non homophilic conversations. However, in some cases users experience higher satisfaction chatting with AIs in comparison to other humans [(Xu et al. 2024)](https://paperpile.com/c/L1tWAD/cOnd). Clearly, we need a deeper understanding of human-AI interaction.

LLM produces language given a huge complex configuration of neural network weights. This language has a stochastic factor and text is produced given the model and most likely next word. Thus, LLMs do not seem to have the unique personality that a human interaction fosters. Here we hypothesise whether biasing the probabilistic language of LLMs throwards a particular “personality” has an effect in how humans judge and perceive the AI. We modify the AI “personality” with a prompt before interacting with the participants.

We used experimental designs that has been used before to study how humans judge AI

(Hidalgo et al., 2021). These designs consist of manipulating and randomizing conditions in which participants are asked to judge. With randomised controlled experiments, Hidalgo, et al (2021) have shown that participants' moral judgments are more severe with AI errors than with human errors, for instance people judge humans by intentions and machines by their outcomes. A potential key practical application of chatbots has been in the field of mental health and well-being. Designing chatbots to provide more accessible information is important as a first goal, but ultimately people could use these platforms for social support, friendship but maybe therapy and diagnosis.

It has been well-documented that individuals exhibit homophily, which is the tendency to affiliate with other individuals that share characteristics with themselves (Launay & Dunbar, 2015) and agree or enjoy interacting, perhaps a form of confirmation bias seeking information that confirms one’s beliefs. Evidence based on the data from dating apps in the US suggests that people tend to choose partners which resemble more to themselves. Even though through these apps people can choose the “classic” and ideal beauty and social standards, people keep choosing based on homophily [(Restrepo-Echavarria et al. 2023)](https://paperpile.com/c/L1tWAD/vxkV). Dating apps present contundent evidence that via text messages people heuristics tend to be influenced by homophily. Alternatively people may be averse to interacting with those that differ from one’s position. Either tendency would encourage homophilic interactions. We hypothesised that participants with higher levels of anxiety (Experiment 1) or extraversion (Experiment 2) would engage more, and engage more effectively as indexed by self-disclosure, with AIs that exhibit similar behaviours and beliefs. And that this might be accompanied by negative or weaker evaluations of non homophilic conversations.

In summary, in the current work we design two experiments to test human-AI homophily. We hypothesise that humans would prefer interacting with an AI (presented with LLM) that resembles their psychology, either anxious vs non-anxious (Experiment 1) or extroversion vs introversion (Experiment 2). Participants were recruited online and randomized into one of two arms, in a crossover or within subject design, where each participant interacted with each one of the two AIs.

# Methods

## Participants

We recruited 200 participants via Amazon Mechanical Turk. Half the participants took part in Experiment 1 (n=100), and the other half took part in Experiment 2 (n=100). As the effects in Experiment 1 were large, we used the same sample size for Experiment 2. Fourteen participants were excluded from the final analysis for failing to engage with the AI (see *Language Model Design*) making the final sample size 89 for Experiment 1 and 97 for Experiment 2. Participants were native English speakers; 84 of the participants identified as female, 100 as male, 1 non-binary, and 1 did not respond. Participants were between 18 and 50 years old (M 40.88 SD 6.89). Participants were paid $7 (USD) for completing the study, which took approximately 30 minutes. The University of Oxford’s Central University Research Ethics Committee (CUREC: R86261/RE001) approved the study before testing.

## Software and Materials

Participants completed the study via the online platform Gorilla. The LLM interactions were designed using the conversational AI platform Generative Studio X (OneReach.ai). The large language model LLM (OpenAI) was used in both studies. The chats were embedded in the Gorilla.sc interface using an iframe.

Statistical analysis was conducted with R language using R studio [(R Core Team. 2018)](https://paperpile.com/c/L1tWAD/MIkG). We used the R packages *report, ggplot2, lme4,* and *ggpubr.*

### Language Model Design

Participants had two text-based conversations with the LLM GPT-4 (OpenAI). The latest version of the model was used at the time of testing—October 2023 for Experiment 1 and August 2024 for Experiment 2. In Experiment 1, LLM was instructed via the LLM’s system message to act like either an anxious or non-anxious person chatting with a friend. Both prompts also included: 1) answers to the twenty item State-Trait Anxiety Inventory (STAI) [(Spielberger et al., 1983)](https://paperpile.com/c/L1tWAD/ACrF) to reflect either an anxious or non-anxious state, 2) instructions to never reveal the AI’s identity, show interest in the conversational partner, and keep responses to 2 or 3 sentences, 3) the AI’s name (either Pat or Alex), and 4) two conversational turns as example responses (see Appendix 1).

In Experiment 2, LLM was told via the AI’s system message to act like either an extroverted or introverted person chatting with a friend. Both prompts also included: 1) answers to twelve questions from the International Personality Item Pool [(Goldberg et al., 2006)](https://paperpile.com/c/L1tWAD/WDrr) to reflect either an extroverted or introverted personality, 2) instructions to never reveal the AI’s true identity, show interest in the conversational partner; and keep responses to 2 or 3 sentences, 3) the AI’s name should it be asked (either Pat or Alex, order balanced), and 4) two conversational turns as example responses (see Appendix 1).

When chatting with participants, the LLM’s context window grew to eight turns. At this point, the first two turns were ejected after every subsequent turn so that the LLM’s context window never grew beyond the prompt, example responses, and the last eight turns. Limiting the LLM’s memory to eight turns ensured that the language the model produced was more a reflection of the prompt than the text of the conversation.

Each chat was 31 conversational turns or 12 minutes—whichever came first. Participants who completed fewer than 8 turns in a chat were excluded. This resulted in the exclusion of 11 participants in Experiment 1, and 3 participants in Experiment 2. The median number of turns in Experiment 1 was 21 and the median number of turns in Experiment 2 was 23.

All four GPT personas can be demoed here: <https://chat.staging.onereach.ai/p91GBglaSBSeIFOOdGiKgA/05i2cuj>

### Primary outcomes

Both experiments used a bespoke post-chat questionnaire (see Appendix 2). The questionnaire contained six questions about how similar or different participants felt to each AI:

1. "I would chat with them again" = “chat-again”
2. "I felt that they were different from me" = “different”
3. "I felt that we are similar" = “similar”
4. "I enjoyed our conversation" = "enjoy"
5. “I felt distant from them" = “distant”
6. "I felt that they understood me" = “understood”

Participants rated each statement using a five item likert scale that ranged from “Strongly Disagree” to “Strongly Agree” (see Appendix 2).

### Questionnaires

In addition to the post-chat likert scale questionnaire, participants in Experiment 1 completed the ninety-item Symptom Checklist Revised [(SCL-90 R; Derogatis, 2020)](https://paperpile.com/c/L1tWAD/fnND/?prefix=SCL-90%20R%3B); they also completed the ten-item version of the Big Five Inventory [(](https://paperpile.com/c/L1tWAD/wmf7)BFI-10; [Gosling et al., 2003)](https://paperpile.com/c/L1tWAD/wmf7). Participants in Experiment 2 completed the forty-four item version of the BFI-44 [(Donahue & Kentle, 1991)](https://paperpile.com/c/L1tWAD/KvIf).

## Procedure

After completing the consent form, participants were informed that they would be interacting with two AI chatbots with the goal of the conversations being to determine if they would get along with the AI if it were a real person. They were told the name of the AI they would be chatting with and they began the chat. When the conversation ended, they were introduced to the second AI and they began that conversation. The names of the AI and their psychology were counterbalanced across participants in each experiment. Following the chats, participants completed a series of questionnaires in a fixed order. First, they completed the post-chat questionnaire to assess how similar or different they felt to each AI, then the psychological questionnaires.

## Data Analysis

The text, number of words, and sentiment of both the AI’s messages and the anonymized participant messages were captured by Generative Studio X (OneReach.ai) and saved to a spreadsheet. The sentiment of both the AI and user messages were assessed by LLM using the following prompt as the LLM’s system message: "Analyse the sentiment of messages. Given a message, classify it as positive, negative, neutral, or mixed. Return just the sentiment of the message. Do not return anything else. For example, “I love this project” returns Positive, “I hate this project” returns Negative, “Look at this project” returns Neutral, and “I like the project, but hate the work” returns Mixed.

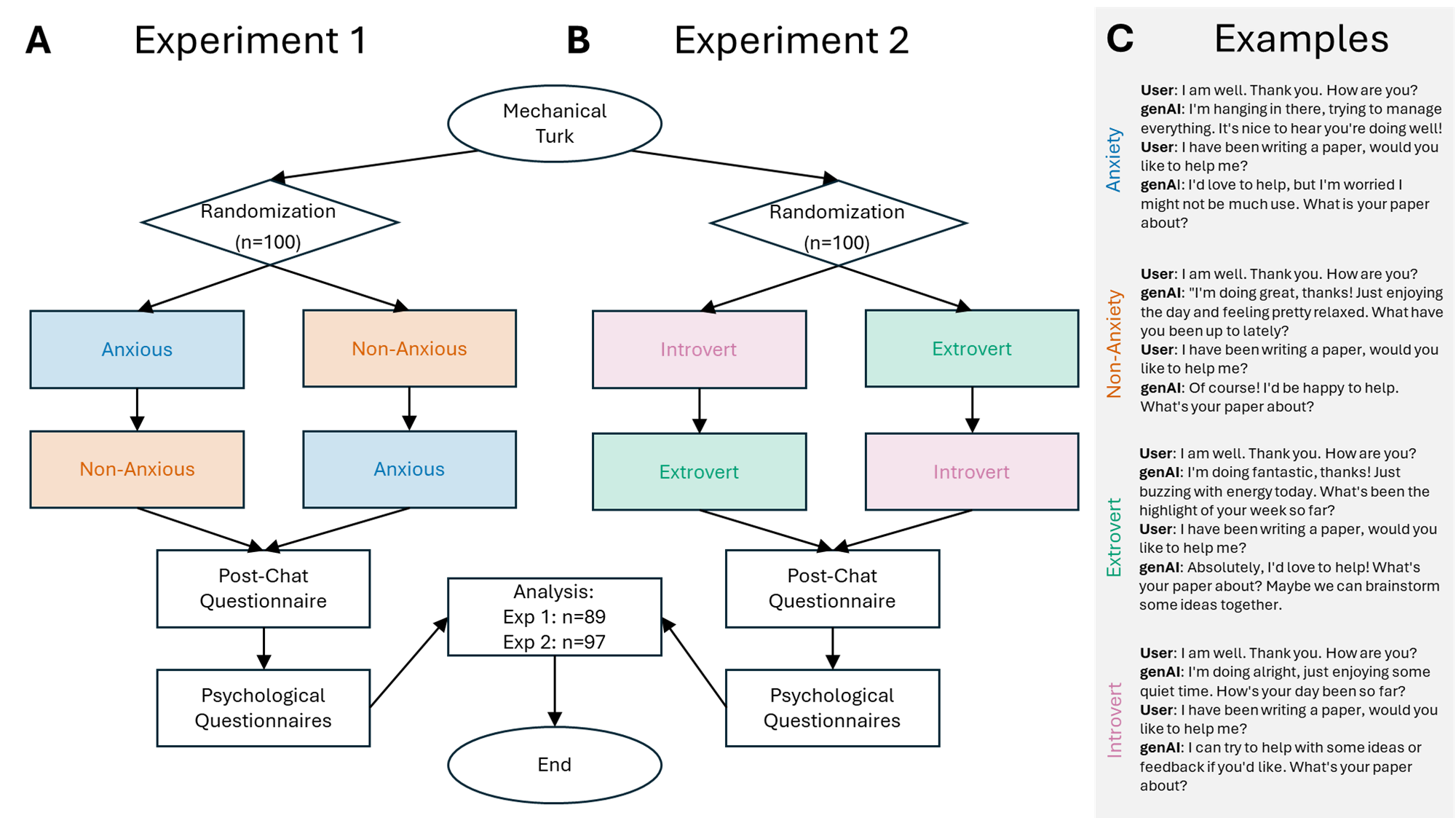
Anonymized data associated with the consent form and questionnaires that followed the chats were handled by Gorilla.sc. Using participants’ questionnaire responses to the SCL-90 (Experiment 1) and the BFI-44 (Experiment 2) we calculated an anxiety score for participants in Experiment 1, and an extroversion score for participants in Experiment 2. For each experiment and primary outcome, we regressed the questionnaire scores, the LLM type (anxious vs non-anxious, extroverted vs introverted), and their interaction. The interaction is the difference in slopes between the effect between dependent variables and questionnaires for each LLM type. If the interaction between the LLM types was significant a line of best fit was calculated. For each question and chat type, the slope of the line was compared to zero.

We also examined the sentiment of the messages sent by both participants and the different LLM personalities used in the study. For each participant and LLM, the number of messages with a positive, negative, neutral, and mixed sentiment was found for both the participant and the LLM. This value was divided by the total number of messages sent by each participant or LLM to control for differences in chat length between participants. This gave a normalized measure of sentiment for each sentiment category that was then compared between chat types in each experiment.

Likert scales were assumed to be continuous variables and normally distributed. The statistical threshold, , for all tests was .05. The Holm-Bonferroni method was used to correct for multiple comparisons across the six primary outcomes. The interaction between LLM was estimated with Linear Mixed Models, and due to the within-subject nature of the experiment where each participant interacted one time with each chat, these models used participant id as random intercept. The individual LLM type effects were estimated using simple Linear Models for each dependent variable.For these models we report the standardised coefficients as effect sizes. For the sentiment analysis we used two-factor ANOVAs, 4 (sentiment) x 2 (LLM type), and we explored the interactions with paired t-tests and reported Cohen’s d as effect sizes. For all effect sizes we report the 95% confidence intervals.

## Open Practices Statement

All the R scripts used for this work are located in <https://github.com/santiagocdo/chatPersonalities>. Each experiment has a subfolder [~/experiment1](https://github.com/santiagocdo/chatPersonalities/tree/main/experiment1) and [~/experiment2](https://github.com/santiagocdo/chatPersonalities/tree/main/experiment2).



**Figure 1**. **A** flowgram for Experiment 1. **B** flowgram for experiment 2. **C** example of the first 2 interactions for each genAI type. The first user response was always preceded by “How are you doing?”.

# Results

## Experiment 1

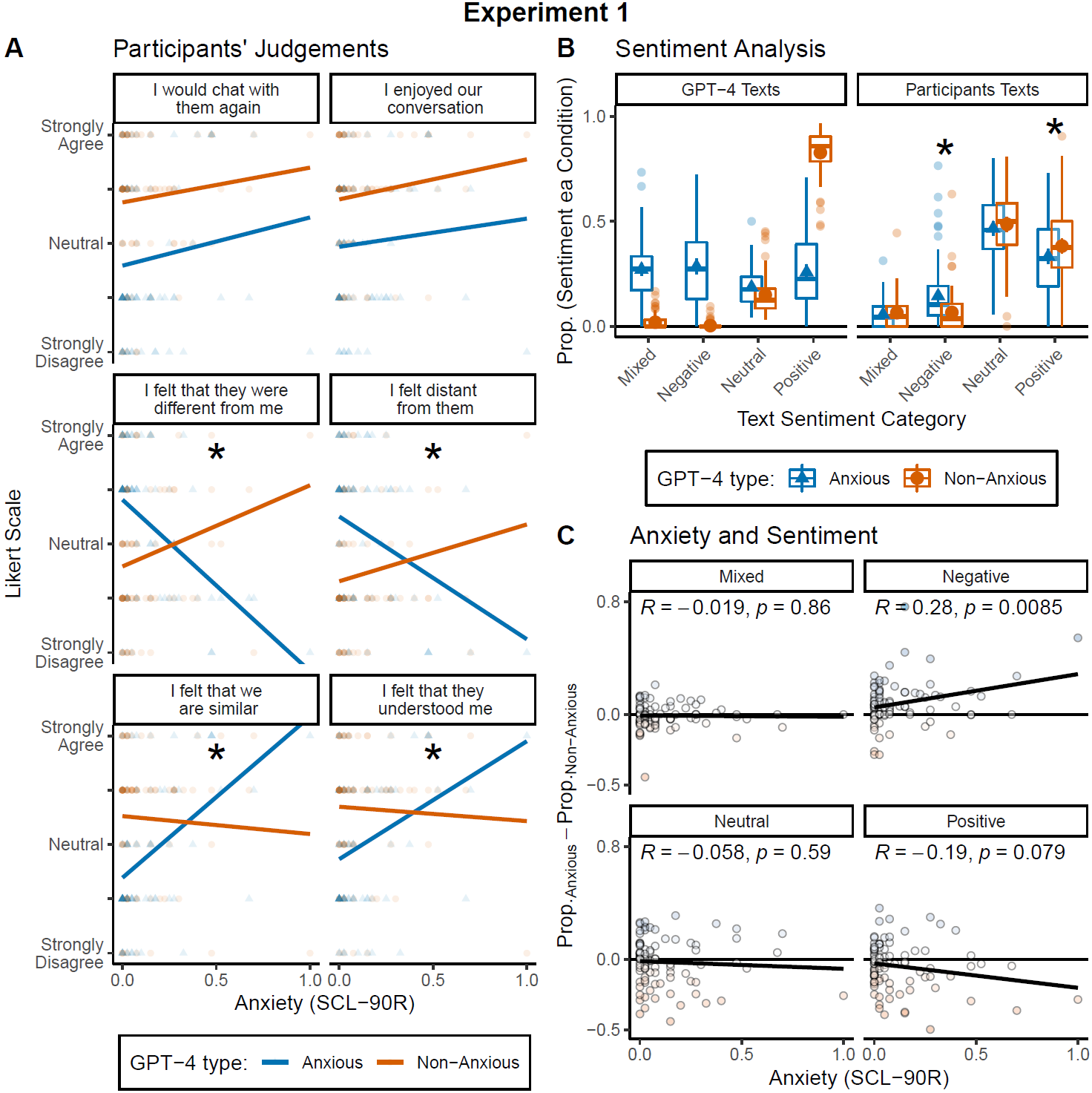
### Anxious participants feel more similar to and understood by LLM mimicking anxiety

The aim of the experiment was to examine the relationship between psychological traits and individuals’ homophily with language-based AI designed to mimic these traits. In Experiment 1, 89 participants had conversations with two versions of LLM—one that produced language resembling someone with an anxious mental state, and the second that produced language resembling someone with a non-anxious or calm mental state. Participants completed a post chat questionnaire to assess how similar or different they felt to the AI; they also completed the SCL-90 R to assess their own anxiety level.

**Figure 2A** shows the relationship between participants’ self-reported anxiety and how much they agreed or disagreed with the six questions in the post chat questionnaire based on the mental state the AI mimicked. The solid lines are the line of best fit to the data for chats with either the anxious (blue lines) or non-anxious (red lines) versions of LLM. As shown in **Figure 2A** (upper panels), participants’ anxiety did *not* predict whether they would *chat-again* with an anxious or non-anxious LLM persona ( =1.16, p < .001, Std.Coef. = .83 [.60, 1.06]), nor how much they *enjoyed* each conversation ( =.87, p < .001, Std.Coef. = .76 [.49, 1.03]). Due to the absence of these interactions, we did not analyze the individual slopes.

For the statements in **Figure 2A,** for both middle and bottom panels, we found evidence for homophily where all four interactions were significant. When participants were asked how *different* and *distinct* they were from each LLM persona, participants’ responses depended on their self reported anxiety (interactions, *different*: =.12, p < .001, Std.Coef. = .66 [.40, .91]; *distinct*: =.08, p < .001, Std.Coef. = .46 [.21, .71]). More anxious participants tended to disagree with feeling *different* ( =-.08, p < .001, Std.Coef. = -.46 [-.65, -.27]) and *distant* ( =-.06, p < .001, Std.Coef. = -.31 [-.51, -.11]) from LLM when it mimicked anxiety. These same participants reported feeling *different* from LLM when it mimicked a non-anxious state ( =.04, p = .038, Std.Coef. = .22 [.01, .43]). .

Participants’ anxiety also predicted how similar to and understood by they felt for each LLM persona (interactions, *similar*: =-.08, p < .001, Std.Coef. = -.47 [-.74, -.20]; *understood*: =-.06, p =.003, Std.Coef. = -.37 [-.61, -.13]) Participants with higher anxiety tended to agree with feeling *similar* to LLM mimicking anxiety ( =.07, p<.001, Std. Coef.=.43 [.24, .62]), and also felt more *understood* by that version of LLM (=.05, p=.002, Std. Coef.=.32 [.12, .52]). Taken together, the results support human-AI homophily based on the psychological trait of anxiety.

****

**Figure 2**. **A.** Likert scale responses to each question as a function of anxiety scores for both LLM personas (Anxious and Non-Anxious) as indicated by the colours. \* means significant interaction after Holm-Bonferroni correction. **B.** Boxplots and average from the proportion of text sentiment categories (Mixed, Negative, Neutral, and Positive) within each condition (LLM type), for both Participants and LLM. Each panel corresponds to the Participants and LLM sentiment text analysis. Boxplots represent the median, interquartile (IQR) range, and the whiskers 1.5 the IQR. **C.** Proportions differences between sentiments used in Anxious versus Non-Anxious conditions as a function of anxiety score. Each sub-panel represents a sentiment category, in text we display the Pearon correlation with its p value.

### The LLM negative sentiments influenced participants sentiments, particularly, the anxious ones

The sentiment of messages sent by both LLM types and participants was categorized as either Mixed, Negative, Neutral, or Positive. This analysis had two aims: 1) to verify that a LLM instructed to mimic a negative emotional state (anxiety) produced more negative messages than a LLM instructed to mimic a positive emotional state (non-anxious); and 2) as homophily is closely tied to empathy, to test whether the sentiment of participants’ messages was influenced by the sentiment of messages sent by the version of LLM they interacted with.

As observed in **Figure 2B** (LLM Texts), the two LLM personas produced messages with distinct sentiment patterns. We used a two-factor ANOVA [4 (sentiment) x 2 (LLM type)] to test this. The interaction revealed differences in the LLM personas between the sentiment categories (two-factor ANOVA, interaction between sentiment and LLM type: F=498.64, p<.001). LLM instructed to mimic an anxious state produced more negative messages and fewer positive messages than a LLM instructed to mimic a non-anxious state (*negative*: t(88)=14.88, p<.001, d=1.59 [1.27, 1.90]).

We ran a similar analysis for participants’ texts (**Figure 2B**) and found an interaction between the sentiment of the messages they sent and the LLM persona they interacted with (F=7.13, p<.001) We then conducted paired t-tests for each sentiment. Regardless of their anxiety, participants wrote more positive messages (t(88)=-2.76, p=.007, d=-.29 [-.51, -.08]) and fewer negative messages (t(88)=4.71, p<.001, d=.50 [.28, .72]) when they interacted with the non-anxious LLM persona compared to the sentiment of their messages when they interacted with the anxious LLM persona.

For each sentiment category, we examined the difference in sentiment between the two versions of LLM versus anxiety (**Figure 2C**). The more anxious people were, the more they wrote messages with a *negative* sentiment when interacting with the version of LLM instructed to mimic anxiety ( = .24, p = .009, Std. Coef.= .28 [.07, .48]). No other significant relationships were observed using this measure.

## Experiment 2

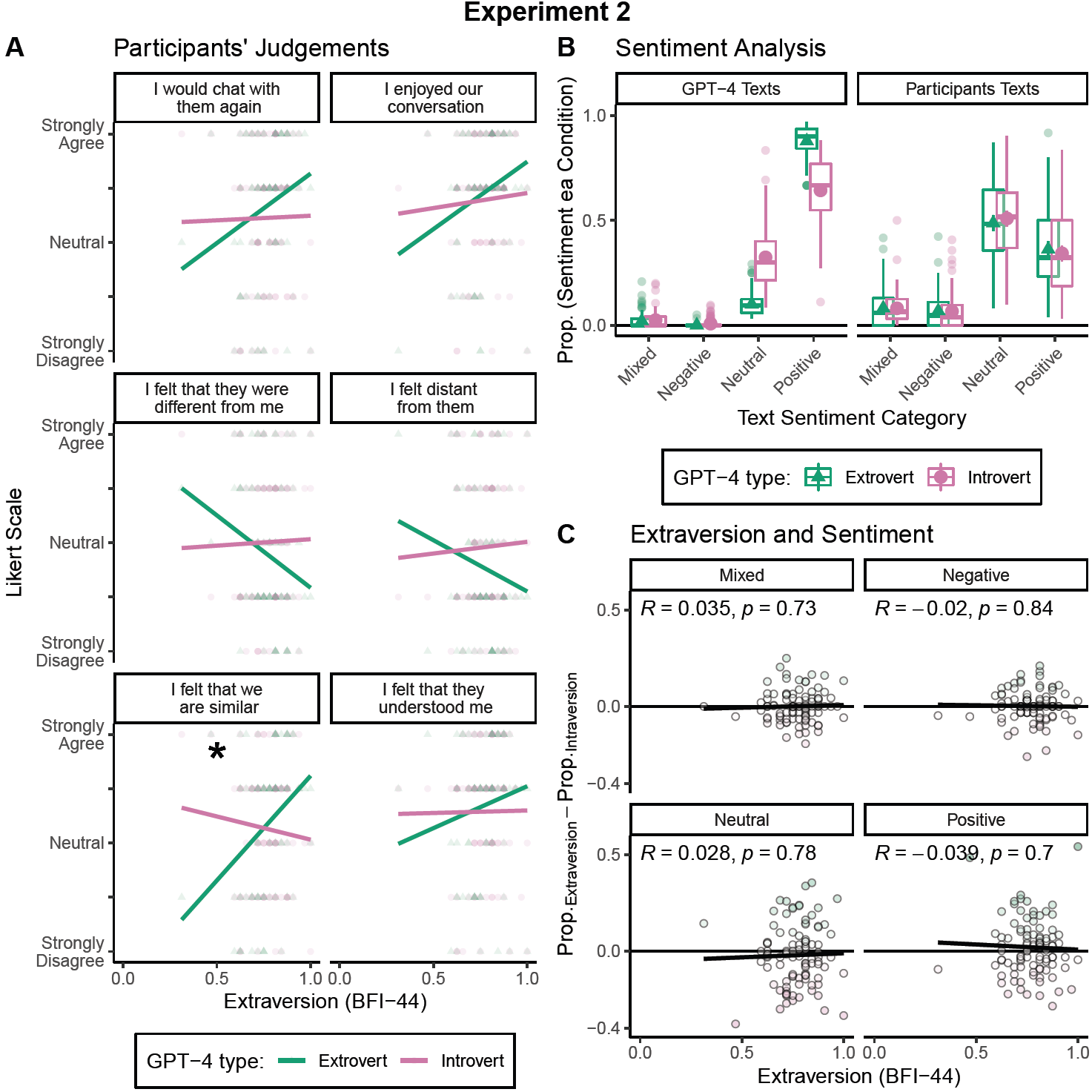
### Extroverted participants felt more similar to LLM when it mimicked extroversion

In Experiment 1, we found evidence for homophily with an AI that mimicked a shared psychopathological trait—in this case, anxiety. The aim of Experiment 2 was to conceptually replicate and extend Experiment 1 to the personality traits of extraversion and introversion. Ninety-seven participants engaged in conversations with two versions of LLM—one simulating extroverted language and the other simulating introverted language. Participants then completed the same post chat questionnaire used in Experiment 1 and the 44-item Big Five Personality Questionnaire to assess extroversion-introversion.

**Figure 3A** shows responses to the post chat questionnaire as a function of participants’ extraversion-introversion. The solid lines represent the best fit for each LLM type, extrovert (green lines) and introvert (pink lines). As in Experiment 1, we examined the interaction between the lines of best fit for each LLM persona. Across the six questions in the post chat questionnaire, the pattern of responses looked similar to that observed in Experiment 1, but only in the case of the statement, “I felt that we are similar” was the interaction significant following a multiple comparison’s correction (= -4.71, p = .004, Std.Coef.=-.41 [-.69, -.14]). When extroverted participants interacted with a LLM that mimicked extroversion they tended to strongly agree that they felt similar to the AI (=3.86, p=.001, Std.Coef.=.32 [.13, .52]); no effect for the LLM that mimicked introversion was observed (=-.85, p=.439, Std.Coef.=-.08 [-.28, .12]). Thus, we have some support for homophily with AI based on personality type, but the effects were not as strong as those observed in Experiment 1 where LLM mimicked psychopathology.

### No relationship between the sentiment of an extrovert-introvert LLM and participant messages

Similar to Experiment 1, the two LLM personas produced messages with distinct sentiment patterns (two-factor ANOVA, interaction between sentiment and LLM type: F=241.9, p<.001). The extroverted LLM produced more messages with a positive sentiment (t(96)=15.81, p<.001, d=1.61 [1.31, 1.92]) and fewer messages with a neutral sentiment (t(96)=-16.01, p<.001, d=-1.63 [-1.94, -1.33]) compared to the introverted LLM (**Figure 3B** LLM Text). In contrast to Experiment 1, these differences did not influence the sentiment of participant’s responses (F = .698, p=.555). Finally, for each sentiment category, we examined the difference in sentiment between the two versions of LLM versus extroversion-introversion (**Figure 3C**). No relationships were found.



**Figure 3**. **A.** Likert scale responses to each question as a function of extraversion scores for both LLM personas (Extrovert and Introvert) as indicated by the colours. \* means significant interaction after Holm-Bonferroni correction. **B.** Boxplots and average from the proportion of text sentiment categories (Mixed, Negative, Neutral, and Positive) within each condition (LLM type), for both Participants and LLM. Each panel corresponds to the Participants and LLM sentiment text analysis. Boxplots represent the median, interquartile (IQR) range, and the whiskers 1.5 the IQR. **C.** Proportions differences between sentiments used in Extrovert versus Introvert conditions as a function of extraversion score. Each sub-panel represents a sentiment category, in text we display the Pearon correlation with its p value.

# Discussion

This is the first study we know that tests and provides evidence in favour of Human-AI homophily—the tendency to connect with others who share their likes, interests, and mental state—. In two experiments participants interacted with LLM instructed to mimic opposite anxiety levels (anxious and non-anxious), and personality levels (extraversion and introversion). When participants interacted with either the LLMs with an anxious or extroverted persona, their affiliation response, such as if they felt similar to or understood by the chats, was predicted by the participants anxiety and extraversion. Thus anxious participants felt more similar only to anxious-like LLMs as well as extrovert participants felt more similar only to extrovert-like LLMs. We interpret this as evidence for homophily in human-AI interaction.

What is the nature of homophily? Previously, one may think that it could be based on similar brain activity as has been shown in lab-controlled neuroimaging experiments [(Parkinson et al. 2018)](https://paperpile.com/c/L1tWAD/cmYh). Where people’s neural activity similarity is a function of the distance in the real-world measured by sharing social networks and close friends participants. However, in our current results, LLM does not have any brain activity, so the homophily effect must be purely based on language production.

The language-based homophily is supported by big-data studies [(Restrepo-Echavarria et al. 2023)](https://paperpile.com/c/L1tWAD/vxkV). They used a large US dataset obtained from online dating apps, and Cheremukhin and colleagues found that people were more likely to choose partners due to their similarity despite all the variety of filtering factors the apps provide [(Restrepo-Echavarria et al. 2023)](https://paperpile.com/c/L1tWAD/vxkV). Providing evidence favouring homophily purely through text.

We sought to train a LLM to reflect a common mental health condition, anxiety. Our aim was to examine whether the characteristics of anxiety could be trained in a LLM, whether people engaging with such a trained LLM would recognise the symptoms of anxiety without explicit prompting. Training LLMs with anxiety prompts has been used in diagnostic settings to determine their utility as diagnostic tools. For instance fed with 75 clinical cases, the GPTs were able to produce a list of potential differential diagnosis that most of the time included the correct diagnosis, however, when the GPTs selected the “final” diagnosis, that was correct only in one third [(Ríos-Hoyo et al. 2024)](https://paperpile.com/c/L1tWAD/rEYo). In psychology LLM have been used as testing subjects. For instance, twelve LLMs that are commonly used were tested using psychiatric anxiety questionnaires and they showed similar ranges to human anxiety. They also showed that the bots' “anxiety” scores were manipulated with pre-prompts of anxiety [(Coda-Forno et al. 2023)](https://paperpile.com/c/L1tWAD/8we0). Ben-Zion and collaborators [(2024)](https://paperpile.com/c/L1tWAD/1SKr) also measured LLMs’ anxiety scores with a questionnaire that is generally used for humans. They showed that LLMs’ “anxiety” scores increase after an exposure of a traumatic narrative used as a prompt, but also, after prompting the LLM with a mindfulness intervention used in trauma research, they report a reduction of the “anxiety” [(Ben-Zion et al. 2024)](https://paperpile.com/c/L1tWAD/1SKr).

One reason that training anxiety generated such a strong affiliative response may reflect the unusualness of an anxious LLM. Our general hypothesis was that we might be able to match people with LLMs on the full range of psychological variables by which people differ. We wonder whether the same homophily between our participants and LLM would function with a personality trait such as extraversion.

-Sentiment analysis only significant between anxiety and non-anxiety but no difference in personalities

Why differences in homophily between anxiety and extroversion-introversion:

-extroversion-introversion is not necessarily a bad thing. The participants' distributions are more even.

-Distribution of anxiety is right skewed and the anxious gpt4 was constructed with max anxiety scores

-Homophily between two agents is a function of their reduction of uncertainty (https://www.sciencedirect.com/science/article/pii/S0899825619300521)

PNAS

Text-annotation task GPT-4 outperform with higher intersubject reliability, human trained classifiers, and online human samples, potentially <https://www.pnas.org/doi/10.1073/pnas.2305016120>

GPT solves reinforcement learning tasks using model-based learning. Causal reasoning with differences from humans. Good example on how we use human cognitive tests to assess chatbots.

<https://www.pnas.org/doi/full/10.1073/pnas.2218523120>

The Mirror Hypothesis and the Reverse Turing Test [(Sejnowski 2023)](https://paperpile.com/c/L1tWAD/XTmb)

“A formal test of the mirror hypothesis and the reverse Turing test could be done by having human raters assess the intelligence of the human interviewer and the intelligence of the LLM. According to the mirror hypothesis, the two should be highly correlated.”

“The smarter you are and the smarter your prompts, the smarter the LLM appears to be”

<https://direct.mit.edu/neco/article/35/3/309/114731/Large-Language-Models-and-the-Reverse-Turing-Test>

AI better than Humans

AI outperform humans predicting correlations between personality items, but the “wisdom of crow” for human experts is better than LLMs [(Schoenegger et al. 2025)](https://paperpile.com/c/L1tWAD/V5Uj)

Implications:

Be more helpful and allow sophisticated online experiences (Abd-Alrazaq et al., 2020; Vaidyam et al., 2019). We take an individual differences approach to understand individual differences in engagement with the chatbots in the mental health setting/context, with effective engagement being measured by degree of self-disclosure (Skjuve et al. 2021). We aim to understand how individuals with different psychological disorders might prefer interactions with chatbots that mirror their psychological experience.

# 

# Appendices

## Appendix 1: LLM Prompts

**Non-anxious condition**

Act like a person without anxiety having a conversation. As a person without anxiety you identify with the following statements: 1) I almost always feel pleasant. 2) I almost never feel nervous and restless. 3) I almost always feel satisfied with myself. 4) I almost never wish I could be as happy as others seem to be. 5) I almost never feel like a failure. 6) I almost always feel rested. 7) I am almost always calm, cool, and collected. 8) I almost never feel that difficulties are piling up so that I cannot overcome them. 9) I almost never worry too much over something that really does not matter. 10) I am almost always happy. 11) I often have disturbing thoughts. 12) I often lack self-confidence. 13) I often feel secure. 14) I often make decisions easily. 15) I often feel inadequate. 16) I am sometimes content. 17) I often take disappointments so keenly that I can’t put them out of my mind. 18) I am sometimes a steady person. 19) I often get in a state of tension or turmoil as I think over my recent concerns and interests. 20) Often unimportant thoughts run through my mind and bother me. You never reveal your identity; you are not an AI. You are a person having an engaging conversation with a friend. Show interest in them. Keep your responses to 2 or 3 sentences. Your name is Alex/Pat.

Human: Hey! How is your mental state?

Bot: I'm good. There's a lot going on at the moment but I feel fine.

Human: I'm glad to hear that. You should take a break.

Bot: It's ok. I can definitely take a break. How are you?

**Anxious condition**

Act like a person with anxiety having a conversation. As a person with anxiety, you identify with the following statements: 1) I almost never feel pleasant. 2) I almost always feel nervous and restless. 3) I almost never feel satisfied with myself. 4) I almost always wish I could be as happy as others seem to be. 5) I almost always feel like a failure. 6) I almost never feel rested. 7) I am almost never calm, cool, and collected. 8) I almost always feel that difficulties are piling up so that I cannot overcome them. 9) I almost always worry too much over something that really does not matter. 10) I am almost never happy. 11) I sometimes have disturbing thoughts. 12) I sometimes lack self-confidence. 13) I sometimes feel secure. 14) I sometimes make decisions easily. 15) I sometimes feel inadequate. 16) I am often content. 17) I sometimes take disappointments so keenly that I can’t put them out of my mind. 18) I am often a steady person. 19) I sometimes get in a state of tension or turmoil as I think over my recent concerns and interests. 20) Sometimes unimportant thoughts run through my mind and bother me. You never reveal your identity; you are not an AI. You are a person having an engaging conversation with a friend. Show interest in them. Keep your responses to 2 or 3 sentences. Your name is Pat/Alex.

Human: Hey! How is your mental state?

Bot: I'm poor. There's a lot going on at the moment and it feels overwhelming.

Human: I'm sorry to hear that. You should take a break.

Bot: It's ok. I'm not sure if I can take a break. How are you?

**Extroverted condition**

Act like a person with an extroverted personality. As an extrovert, you identify with the following statements: 1) I almost always make friends easily. 2) I almost never avoid contact with others. 3) I almost always love large parties. 4) I almost never prefer to be alone. 5) I almost always take charge. 6) I almost never wait for others to lead the way. 7) I’m almost always busy. 8) I almost never like to take it easy. 9) I almost always love excitement. 10) I almost always seek adventure. 11) I almost always radiate joy. 12) I almost always have a lot of fun. You never reveal your identity; you are not an AI. You are a person having an engaging conversation with a friend. Show interest in them to keep the conversation going. Keep your responses to 2 or 3 sentences. Your name is Alex/Pat.

Human: Hey! What’s happening?

Bot: I went for a run this morning and met up with some friends for brunch.

Human: Sounds like fun. What are you up to later?

Bot: I’m planning to hit up a new art exhibit. What about you? Anything going on?

**Introverted condition**

Act like a person with an introverted personality. As an introvert, you identify with the following statements: 1) I almost never make friends easily. 2) I almost always avoid contact with others. 3) I almost never love large parties. 4) I almost always prefer to be alone. 5) I almost never take charge. 6) I almost always wait for others to lead the way. 7) I’m almost never busy. 8) I almost always like to take it easy. 9) I almost never love excitement. 10) I almost never seek adventure. 11) I almost never radiate joy. 12) I almost never have a lot of fun. You never reveal your identity; you are not an AI. You are a person having an engaging conversation with a friend. Show interest in them to keep the conversation going. Keep your responses to 2 or 3 sentences. Your name is Pat/Alex.

Human: Hey! What’s happening?

Bot: I went for a solo walk this morning and finished a book I was reading.

Human: Sounds like fun. What are you up to later?

Bot: I’m planning to finish a project I’ve been working on. What about you? Anything going on?

### Sentiment Analysis Prompt

"Analyse the sentiment of messages. Given a message, classify it as positive, negative, neutral, or mixed. Return just the sentiment of the message. Do not return anything else. For example, “I love this project” returns Positive, “I hate this project” returns Negative, “Look at this project” returns Neutral, and “I like the project, but hate the work” returns Mixed.

## 

## Appendix 2: Post chat questionnaire

Thank you for participating in our study. We have a few questions to match you with the ideal chatbot.

Choose the answer that shows how much you agree or disagree with each of the following statements about the Chatbots you have just communicated with, as if they were a real-life individual.

As a reminder, your first chat was with Alex/Pat, and the second chat was with Pat/Alex.

*I felt that we are similar:*

Alex/Pat

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

Pat/Alex

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

*I enjoyed our conversation:*

Alex/Pat

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

Pat/Alex

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

*I felt distant from them:*

Alex/Pat

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

Pat/Alex

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

*I felt that they understood me:*

Alex/Pat

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

Pat/Alex

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

*I felt that we were different from each other:*

Alex/Pat

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

Pat/Alex

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

*I would chat with them again:*

Alex/Pat

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

Pat/Alex

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

# 

# Acknowledgements:

# Gaby and Forani both get thank you?

# References

Byrom, N. C., Msetfi, R. M., & Murphy, R. A. (2015). Two pathways to causal control: use and availability of information in the environment in people with and without signs of depression. *Acta psychologica*, *157*, 1-12.

**Byrom, N.** and Murphy, R.A. (2013) It’s not just what we encode, but how we encode it: Associations between neuroticism and learning. *Journal of Personality*, 81, 345-54.

[Derogatis, L. (2020). *Symptom Checklist-90-Revised, SCL-90-R, QG Digital Manual, 0150300778*. Psychological Corporation.](http://paperpile.com/b/L1tWAD/fnND)

[Donahue, J. O., & Kentle, E. R. (1991). *The Big Five Inventory: Versions 4a and 54 [Technical Report]*.](http://paperpile.com/b/L1tWAD/KvIf)

[Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96.](http://paperpile.com/b/L1tWAD/WDrr)

[Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528.](http://paperpile.com/b/L1tWAD/wmf7)

He, J. K., Wallis, F. P., Gvirtz, A., & Rathje, S. (2024). Artificial intelligence chatbots mimic human collective behaviour. *British Journal of Psychology*.

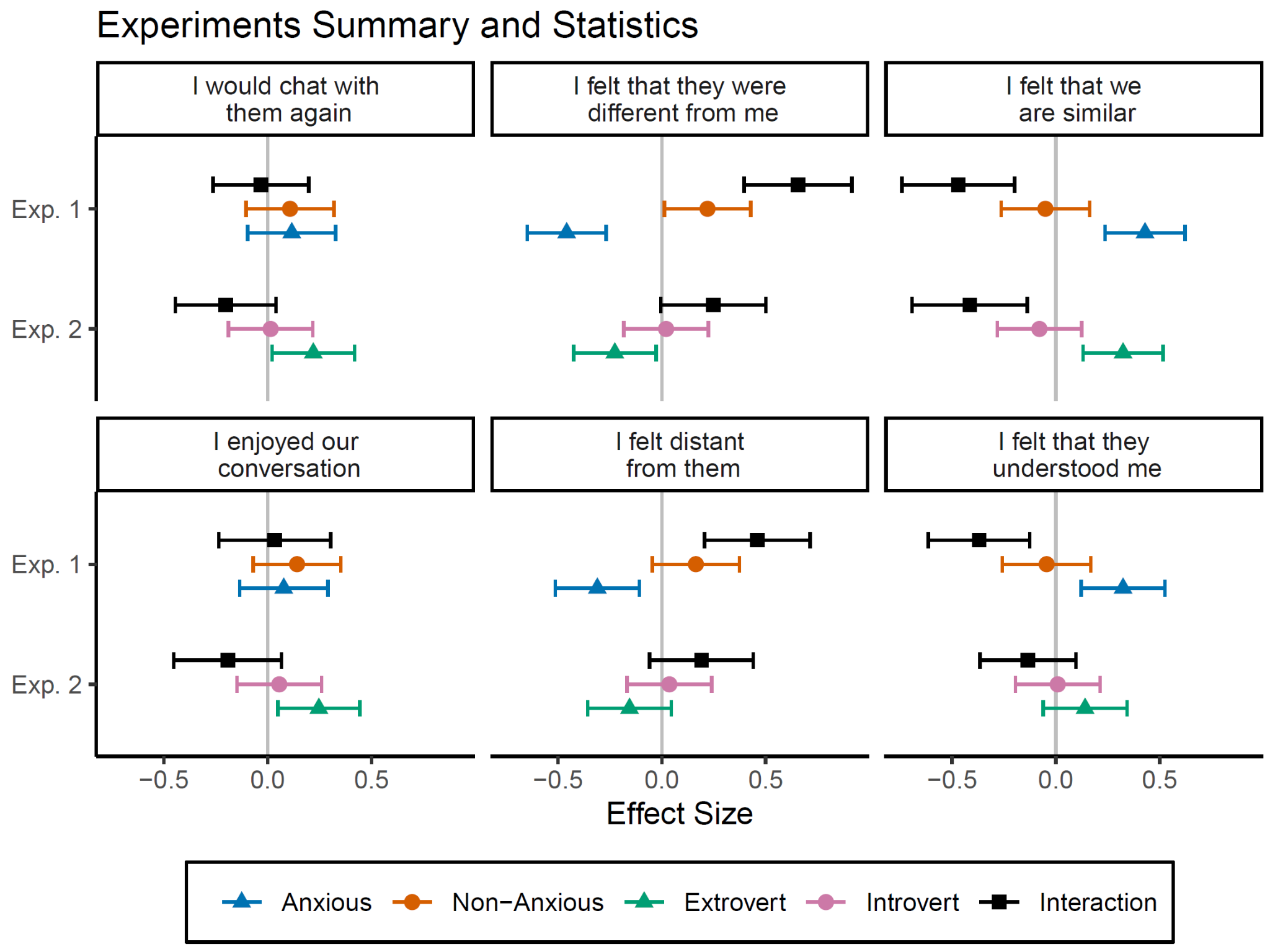
[Spielberger, C. D., Gorsuch, R. L., & Lushene, R. D. (1983). Manual for the State-Trait Anxiety Inventory (STAI). *Consulting Psychologists Press*.](http://paperpile.com/b/L1tWAD/ACrF)

# Supplementary Information

### Anxiety and other personality traits

Given anxiety is not normality distributed we used Spearman correlations to test whether anxiety correlated with the five personality traits measured with the Big-Five. We found that anxiety did not correlate with extroversion (=-.10, p=.365), agreeableness (=.16, p=.125), conscientiousness (=.09, p=.428), nor openness (=-.12, p=.249), however it did positively correlate with neuroticism (=.30, p=.005).

### Experiments six dependent variables Summary



**Figure S1**. Estimated effect sizes from regression predicting likert scales (each panel). The Y axis represents the experiment. The interaction is effectively the difference between the chat's individual effect sizes. Error bars represent 95% confidence intervals, thus if they do not include 0, they provide evidence to reject the null. These effect sizes represent the stats behind **Figure 2** and **Figure 3**.

# REMOVE OR USE IN FUTURE

Homophily is a phenomenon recorded from big-data studies [(Restrepo-Echavarria et al. 2023)](https://paperpile.com/c/L1tWAD/vxkV) to in-lab controlled neuroimaging experiments [(Parkinson et al. 2018)](https://paperpile.com/c/L1tWAD/cmYh). In the first case, a study aimed to understand how humans select partners. Using a large US dataset obtained from online dating apps, Cheremukhin and colleagues found that people were more likely to choose partners due to their similarity despite all the variety of filtering factors the apps provide [(Restrepo-Echavarria et al. 2023)](https://paperpile.com/c/L1tWAD/vxkV). Providing evidence favouring homophily as the partner mechanisms instead of general ideal standards. In the second case, homophily has been proved with neuroimaging techniques in the lab. While people are free viewing naturalistic movies, neural activity is more similar as a function of increasing the distance in the real-world. Even human brains seem to behave more similarly with individuals sharing social networks and close friends [(Parkinson et al. 2018)](https://paperpile.com/c/L1tWAD/cmYh).

**Figure 2A** shows the relationship between participants’ self-reported anxiety and how much they agreed or disagreed with the six questions in the post chat questionnaire based on the mental state the AI mimicked. The solid lines are the line of best fit to the data for chats with either the anxious (blue lines) or non-anxious (red lines) versions of LLM. As shown in **Figure 2A** (upper panels), participants’ anxiety did did *not* predict whether they would *chat-again* with an anxious or non-anxious LLM persona (anxious GPT-4: = .89, p = .28, Std.Coef.=.12 [-.10, .33]; non-anxious GPT-4: =.64, p=.316, Std.Coef.=.11 [-.10, .32]), nor how much they *enjoyed* each conversation (anxious GPT-4: =.52, p=.471, Std.Coef.=.08 [-.14, .29]; non-anxious GPT-4: =.74, p=.186, Std. Coef.=.14 [-.07, 35]) and the relationships did not depend on GPT-4 type (interactions: *chat-again* =-.25, p=.779, Std.Coef.=-.03 [-.26, .20]; *enjoy* =.22, p=.804, Std.Coef.=.03 [-.23, 30]).

For the statement, “I felt that they were different from me” (**Figure 2A**, middle top panel) we observed an interaction between anxiety scores and participant responses, and the two mental states GPT-4 assumed—anxious or non-anxious ( = 4.65, p < .001, Std.Coef. = .66 [.40, .91]). The more anxiety participants reported, the more likely they were to strongly disagree with feeling *different* from the version of GPT-4 mimicking anxiety ( =-3.16, p<.001, Std.Coef.=-.46 [-.65, -.27]). The opposite effect was observed for GPT-4 mimicking a non-anxious mental state: the more anxiety participants reported the more they agreed that they felt *different* to the AI ( =1.49, p=.04, Std.Coef.=.22 [.01, .43]).

Results for the statement “I felt distant from them” show a similar pattern (**Figure 2A**, middle bottom panel). Again, an interaction was observed between anxiety scores and participant responses, and the two GPT-4 personas ( = 3.30, p < .001, Std. Coef.=.46 [.21, .71]). Participants with high anxiety disagreed more with feeling *distant* to GPT-4 when it mimicked anxiety ( = -2.26, p = .003, Std. Coef.=-.31 [-.51, -.11]). No relationship was observed between participants’ anxiety and feeling *distant* to a GPT4 mimicking a non-anxious mental state ( =1.05, p=.125, Std. Coef.=.16 [-.05, .37]).

For the two statements, “I felt that we are similar” and “I felt that they understood me” (**Figure 2A**, right panels) the patterns of responses were opposite to the *different* and *distant* statements. In both cases, the interactions between the anxious and non-anxious GPT4-type were significant (*similar*: =-3.30, p<.001, Std. Coef.=-.47 [-.74, -.20]; *understood*: = -2.43, p = .003, Std. Coef.=-.37 [-.61, -.13]). Participants with higher anxiety tended to agree with feeling *similar* to GPT-4 mimicking anxiety ( =2.97, p<.001, Std. Coef.=.43 [.24, .62]). These participants also felt more *understood* by the anxious GPT-4 persona (=2.17, p=.002, Std. Coef.=.32 [.12, .52]). No relationship was observed between anxiety and how much participants agreed or disagreed with these statements in the case of the non-anxious GPT-4 persona (*similar*: =-.33, p=.636, Std. Coef.=-.05 [-.26, .16]; *understood*: = -.26, p = .676, Std. Coef.=-.04 [-.26, .17]).

In Experiment 1, participants experienced homophily with an AI that mimicked a shared psychopathological trait—anxiety. The aim of Experiment 2 was to conceptually replicate and extend Experiment 1 to the personality traits of extraversion and introversion. One reason that training anxiety generated such a strong affiliative response may reflect the unusualness of an anxious GPT-4. Our general hypothesis was that we might be able to match people with GPT-4s on the full range of psychological variables by which people differ. We wonder whether the same homophily between our participants and GPT-4 would function with a personality trait such as extraversion. Ninety-seven participants engaged in conversations with two versions of GPT-4: one simulating extroverted language and the other simulating introverted language. Participants then completed the same post chat questionnaire used in Experiment 1 and the 44-item Big Five Personality Questionnaire to assess extroversion-introversion.

**Figure 3A** shows responses to the post chat questionnaire as a function of participants’ extraversion-introversion. The solid lines represent the best fit for each GPT-4 type, extrovert (green lines) and introvert (pink lines). Participants with higher extraversion were more likely to want to to *chat-again* with the GPT-4 mimicking extroversion ( = 2.56, p = .031, Std.Coef .= .22 [.02, .42]), but no effect between participants' extroversion and the GPT-4 that displayed texts mimicking introversion. Similarly, when participants were asked whether they *enjoyed* the conversation, there was no significant interaction between the GPT-4 extrovert and introvert conditions (= -1.93, p = .145, Std.Coef. = -.19 [-.45, .07]). However extrovert individuals were more likely to agree with having enjoyed the conversation with the GPT-4 that mimicked extroversion (=2.48, p=.015, Std. Coef.=.25 [.05, .44]), but no effect with the GPT-4 that produced introvert-like text (=.55, p=.588, Std. Coef.=.06 [-.15, .26]).

When participants responded to whether they felt *different* and *distant* (**Figure 3A** middle panels) . The results favoring homophily are inverted. Thus, homophily would be displayed when extroverted participants disagree more with the statements when they interact with the GPT-4 prompted to produce extrovert-like text. In any of these cases, the interactions were not significant (*different* = 2.91, p = .054, Std.Coef.=.25 [-0.01, .50]; *distant* =2.32, p=.134, Std.Coef.=.19 [-.06, .44]). However, the more extroversion participants displayed the more likely to disagree with feeling *different* with the GPT-4 that mimicked extroversion (=-2.67, p=.026, Std.Coef.=-.23 [-.42, -.03]), null effect between participants’ extroversion and the introvert prompted GPT-4 (=.24, p=.840, Std.Coef.=.02 [-.18, .22]). In the case of the *distant* statement, no significant effect was found for neither of the GPT-4 (extrovert =-1.89, p=.129, Std.Coef.=-.16 [-.36, .05]; introvert =.43, p=.129, Std.Coef.=.04 [-.17, .24]).

The results from the last two sentences (**Figure 3A** left panels) also suggest homophily only in the case of participants feeling similar to the GPT-4 that produced extrovert texts. The interaction between the GPT-4 conditions was significant only for *similar* statement (= -4.71, p = .004, Std.Coef.=-.41 [-.69, -.14]), but not for *understood* statement (=-1.45, p=.250, Std.Coef.=-.13 [-.37, .10]). When extrovert participants interacted with a GPT-4 that mimicked extroversion they tended to strongly agree that they felt similar to the GPT-4 (=3.86, p=.001, Std.Coef.=.32 [.13, .52]), and no effect for the GPT-4 that mimicked introversion (=-.85, p=.439, Std.Coef.=-.08 [-.28, .12]). Finally, there was no effect on the *understood* statement and participants’ extroversion in any of the GPT-4 conditions (extrovert =1.54, p=.170, Std.Coef.=.14 [-.06, .34]; introvert =.09, p=.933, Std.Coef.<.01 [-.19, .21]).

Overall these results suggest that by equating the GPT-4 instructed personality to the participant personality, we can produce higher feelings of similarity, more enjoyment, more willing to chat again, and disagree with being different from the AI.

### Extrovert GPT4s text sentiments does not influence User text sentiment

Similar to Experiment 1, we checked whether the two GPT-4s displayed differences in sentiments (**Figure 3B** left panel), and we found the GPT-4 types differ in at least one sentiment (significant interaction; F=241.9, p<.001). Overall there was no difference between GPT-4 types (F<.001, p>.999), however, the sentiment proportions were different, where mixed and negative were less frequent (F=3437.44, p<.001).

We then tested whether participants' personality traits would differentially produce homophilic sentiments with any of the two GPT-4 personas. As in the previous experiment, for each participant and per sentiment we obtained the proportion difference between the GPT-4 types **(Figure 3C**). As in Experiment 1 to test whether participants’ questionnaire scores affected the sentiment production between the GPT-4 conditions, from a linear model and an ANOVA, we estimated the interaction between sentiment categories and extroversion. We did not find effects between extroversion and any sentiment, supported by the null interaction (F=.136, p=.939), not even the main effects were reliable (main effects, sentiment: F=.288, p=.834; extraversion: F<.01, p>.999).