Human Enthusiasm and Perceived Capabilities of Large Language Models

Authors: Wu Yueting, Wang Bo, Bao Han Wu Shuang, Li Ruonan, Wu Yi, Wang Jiaqi, Cheng Cheng, Yang Li, Wang Bo, Yang Li

Date: 2025-07-18T00:00:00+00:00

Abstract

With the advancement in capabilities and widespread application of Large Language Models (LLMs), society is gradually transitioning from traditional interpersonal interaction to a multi-level interactive structure that integrates interpersonal, human-computer, and machine-machine interactions. Against the backdrop of increasingly deepened human-LLM interaction, investigating how humans perceive LLMs has become an important research topic. This study systematically examines human perception patterns of LLMs through three studies. Study 1 found that, consistent with the perception of humans, humans primarily perceive LLMs along two dimensions: warmth and competence. However, in general contexts, unlike the warmth primacy in human perception, humans exhibit competence primacy when perceiving LLMs. Study 2 explored the primacy effects of warmth and competence in predicting different attitudes, with results indicating that both warmth and competence positively predict humans' continuous usage intention and likability toward LLMs, where competence demonstrates higher predictive power for continuous usage intention, while warmth demonstrates higher predictive power for likability. Study 3 further explored differences in human perception between LLMs and other humans, revealing that humans' warmth evaluations of LLMs show no significant difference from those of humans, but competence evaluations of LLMs are significantly higher than those of humans. This study provides a theoretical foundation for understanding human perception of LLMs and offers new perspectives for the design optimization of artificial intelligence and research on human-machine collaboration mechanisms.

Full Text

Preamble

Humans Perceive Warmth and Competence in Large Language Models

WU Yueting[1], WANG Bo[2,3], *BAO Hanwushuang*[4], *LI Ruonan*[1], *WU Yi*[1], *WANG Jiaqi*[1], *CHENG Cheng*[3], *YANG Li*[3,5]

[1]School of Education, Tianjin University, Tianjin 300354, China
[2]College of Intelligence and Computing, Tianjin University, Tianjin 300354, China
[3]Institute of Applied Psychology, Tianjin University, Tianjin 300354, China
[4]School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China
[5]Laboratory of Suicidal Behavior Research, Tianjin University, Tianjin 300354, China

Abstract

As Large Language Models (LLMs) continue to advance in capability and proliferate across applications, society is transitioning from traditional interpersonal interactions toward a multilayered structure that integrates human-human, human-machine, and machine-machine interactions. Against this backdrop of deepening human-LLM engagement, understanding how humans perceive LLMs has emerged as a critical research question. This study systematically investigates human perception patterns of LLMs through three empirical studies.

Study 1 reveals that, consistent with human perception of other humans, people primarily perceive LLMs through two fundamental dimensions: warmth and competence. However, in general contexts, unlike the warmth-priority effect observed in human-to-human perception, humans exhibit a competence-priority effect when perceiving LLMs. Study 2 examines the predictive priority of warmth and competence across different attitudinal outcomes. Results demonstrate that both dimensions positively predict continued use intention and liking of LLMs, with competence showing stronger predictive power for continued use intention, while warmth demonstrates greater predictive power for liking. Study 3 further explores perceptual differences between LLMs and other humans, finding no significant difference in warmth ratings between LLMs and humans, but significantly higher competence ratings for LLMs compared to humans.

These findings provide a theoretical foundation for understanding human perception of LLMs and offer new perspectives for AI design optimization and human-machine collaboration research.

Keywords: Large Language Model, Social Cognition, Warmth, Competence

1 Introduction

Rapid advancements in artificial intelligence and natural language processing (NLP) technologies have propelled significant progress in the development and application of Large Language Models (LLMs) [?, ?, ?]. LLMs are driving the realization of artificial general intelligence [?, ?] and have begun exhibiting human-like psychological characteristics. For instance, their Theory of Mind (ToM) capabilities now approach those of six-year-old children [?, ?], they demonstrate performance surpassing human levels on certain tasks [?, ?], display personality traits similar to humans [?, ?, ?], and show capacity to modulate responses based on emotional cues [?, ?].

As LLMs become increasingly integrated into daily life, the scope of human interaction has expanded from traditional self, others, and groups to include LLMs. This transformation profoundly influences social interaction patterns and presents new challenges. In this context, how humans perceive LLMs has become an urgent question: Do humans apply similar theoretical frameworks of social cognition to LLMs as they do to other humans? This question is significant not only for understanding social cognition of intelligent agents but also for providing theoretical insights for AI design optimization and human-machine collaboration mechanisms.

1.1 Dimensions of Human Perception of LLMs

In social life, people continuously perceive and evaluate others, groups, and themselves, with these evaluations directly influencing attitudes, behaviors, and decision-making. Understanding how individuals perceive self and others is therefore fundamental to comprehending human social interaction. Extant theoretical research converges on the notion that interpersonal social evaluation revolves around two primary dimensions, a framework known as the "Big Two" model of social cognition content, validated across numerous cross-cultural studies [?, ?, ?, ?]. Representative theories include the Dual Perspective Model (DPM) and the Stereotype Content Model (SCM). DPM posits that communion and agency constitute core dimensions of self and other perception [?, ?, ?, ?], with communion involving traits such as friendliness and caring, and agency encompassing competence, independence, and subjective initiative. SCM argues that people primarily attend to warmth and competence when perceiving groups [?, ?, ?], where warmth refers to perceived friendliness and kindness, and competence measures task performance and professional quality.

Although DPM and SCM differ in research targets and dimensional connotations, Abele et al. [?, ?] recently proposed an integrated framework through an "adversarial-collaborative" approach, specifying two core dimensions of social cognition content as vertical and horizontal dimensions. The vertical dimension corresponds to competence/agency, while the horizontal dimension corresponds to warmth/communion. Chinese scholars Zuo et al. [?, ?] noted in their review of the Big Two model that "considering communion and agency are easily confused with concepts from personality and other social psychology domains," they recommended using warmth and competence as translations. This study adopts this terminology, using "warmth" and "competence" throughout.

With the widespread application of artificial intelligence, AI agents have gradually integrated into human society as important components of social interaction. Understanding how people perceive AI is therefore key to developing good human-machine interaction. McKee et al. [?, ?] validated that warmth and competence are primary dimensions through which humans perceive traditional AI agents (e.g., voice assistants, recommendation systems, gaming systems, autonomous driving). However, no research has yet validated this theoretical applicability in the LLM domain. Although LLMs remain algorithm-driven, their language generation capabilities blur human-machine boundaries, exhibiting more human-like behaviors. Meanwhile, ethical norms and moral attributes have become increasingly important in LLM development, potentially leading people to attend more to moral attributes when evaluating LLMs. For example, people may question whether LLMs can correctly understand and follow social ethics, which in the Big Two model is typically subsumed under the warmth dimension (morality being a facet of warmth). Additionally, widespread discussions about LLM generative capabilities further highlight their performance on the competence dimension. These considerations suggest that human perception of LLMs likely continues the two-dimensional structure of the Big Two model. Therefore, we propose:

H1: Warmth and competence are the primary dimensions of human perception of LLMs.

1.2 Priority Effects in Dimensions of Human Perception of LLMs in General Contexts

A central question in social cognition content research concerns the relative importance of warmth versus competence, known as the priority effect. Research indicates that in general interpersonal perception contexts, warmth typically precedes competence—individuals evaluating others tend to first attend to warmth before competence, with warmth evaluations eliciting stronger attitudinal and behavioral responses [?, ?, ?, ?]. DPM explains this through the principle of profit maximization [?, ?]. Specifically, in interpersonal interactions, warmth represents other-profitable traits while competence represents self-profitable traits. To maximize self-interest, observers focus more on others' warmth than their competence.

However, in human perception of LLMs, the classification of warmth and competence as other-versus self-profitable does not align perfectly with human-to-human perception. Since LLMs are assumed to lack "self," both their warmth and competence are viewed by human users as serving humans (others) rather than themselves, thus constituting other-profitable traits. Furthermore, compared to warmth, LLM competence typically exhibits more direct reciprocity and practicality, such as providing answers, completing tasks, and assisting decision-making. This efficient return more directly helps users achieve goals, making LLM competence more capable of maximizing human interests. Therefore, we propose:

H2: In human perception of LLMs, competence demonstrates a priority effect over warmth.

1.3 Priority Effects of Perceptual Dimensions in Predicting Attitudes Toward LLMs

As Fiske [?, ?] noted, "thinking is for doing," with research on social cognition content ultimately addressing how it triggers cognitive, emotional, and behavioral responses. The Behaviors from Intergroup Affect and Stereotypes Map (BIAS Map) further explores relationships between perception and emotional/behavioral tendencies [?, ?, ?]. BIAS Map posits that warmth perception influences proactive behavioral tendencies, while competence perception influences passive behavioral tendencies. Based on different combinations of warmth and competence, BIAS Map categorizes social perception targets into four group types and identifies corresponding emotional responses. For example, high-warmth, high-competence groups typically elicit admiration, while high-competence, low-warmth groups elicit envy. Research on traditional AI agents also shows that warmth and competence dimensions predict human attitudes toward them. The Robotic Social Attributes Scale (RoSAS) proposed warmth, competence, and discomfort as three factors describing human perception of robots' social attributes [?, ?], with Scheunemann et al. [?, ?] confirming that warmth and competence are optimal predictors of human-robot interaction preferences.

The predictive power of warmth and competence also reflects priority effect implications. Priority effects involve not only earlier perception of one dimension but also stronger influence of that dimension on behavioral and attitudinal responses [?, ?]. This study therefore addresses two questions: First, can human evaluations of LLM warmth and competence effectively predict attitudinal tendencies toward LLMs? Second, does dimensional priority shift across different attitude predictions? To address these questions, we selected continued use intention and liking as two distinct evaluation targets—typical indicators for assessing human-machine interaction outcomes.

Continued use intention, as a functional cognitive evaluation indicator, reflects the degree to which humans recognize LLMs' task-completion capabilities. It is influenced by several core factors including performance expectancy, effort expectancy, social influence, and facilitating conditions [?, ?], which reflect functional evaluations of LLMs, particularly regarding task performance efficacy and ease of use. Within this framework, human competence evaluations directly influence continued use intention. Liking, as an affective cognitive evaluation indicator, reflects the degree of emotional connection between humans and

LLMs. Unlike continued use intention, liking relates more to humor [?, ?, ?] and anthropomorphism [?, ?, ?]. These factors resemble warmth traits, and LLMs high in warmth typically provide more pleasant interaction experiences, thereby fostering stronger emotional connections. Thus, warmth evaluation plays a more important role in liking formation. Based on this reasoning, we propose:

H3a: Both warmth and competence ratings of LLMs positively predict continued use intention, with competence showing stronger predictive power.

H3b: Both warmth and competence ratings of LLMs positively predict liking, with warmth showing stronger predictive power.

1.4 Comparison of Warmth and Competence Ratings Between Humans and LLMs

While the aforementioned research focuses on exploring perception patterns of LLMs, it remains unclear how LLMs perform on warmth and competence dimensions relative to human perception. Although LLMs have made substantial progress in human-likeness, complex human cognitive, emotional, and behavioral characteristics remain difficult to fully simulate. Specifically, current human evaluations of LLMs' emotional interaction present contradictory features: while existing research shows LLMs can generate empathy responses conforming to social norms [?, ?], human users often describe LLMs as procedural and lacking genuine emotion. Therefore, despite LLMs' capacity to generate socially appropriate emotional expressions, humans may perceive LLMs as inferior to other humans on the warmth dimension due to their lack of emotional depth. In contrast to warmth evaluations, the massive attention LLMs have garnered stems precisely from perceived capability improvements. Research indicates LLMs have surpassed general human levels in multiple domains including logical reasoning [?, ?] and abstract ability [?, ?]. Consequently, we propose:

H4a: Warmth ratings differ between humans and LLMs, with lower warmth ratings for LLMs.

H4b: Competence ratings differ between humans and LLMs, with higher competence ratings for LLMs.

1.5 Overview of Studies

This study aims to construct a theoretical framework for human perception of LLMs, systematically analyzing the dimensions of human perception of LLMs, examining priority effects in general contexts, and investigating whether these effects shift in attitude prediction. Additionally, this study compares perceptual differences between LLMs and other humans. The research fills gaps in human perception of LLMs and provides new theoretical perspectives for AI design and optimization.

2 Study 1: Dimensions and Priority Effects in Human Perception of LLMs

2.1 Study 1a: Free-Response Task

2.1.1 Purpose This study explores the primary dimensions and dimensional priority in human perception of LLMs through a free-response task to test H1 and H2—that warmth and competence are primary dimensions of LLM perception and that competence demonstrates priority over warmth.

2.1.2 Method **Participants.** We recruited 215 participants through an online survey platform. After excluding 8 participants with no LLM usage experience or who failed to follow instructions, the final sample comprised 207 valid participants (124 males, 59.90%; 83 females, 40.10%). All participants volunteered for the study.

To ensure adequate statistical power for the formal study, we conducted power analysis based on pilot data (31 participants). As we planned to use Generalized Linear Mixed Models (GLMM) to test hypotheses, we employed the R package simr [**?**, **?**] for simulation-based power analysis (two-tailed test, $\alpha = 0.05$; 1,000 Monte Carlo simulations). Results showed that with a sample size of 31, power to detect an unstandardized regression coefficient of -0.50 for warmth and competence coverage rates was 88.70% and 89.80%, respectively; power reached 100% for detecting coefficients of -1.00. Thus, the formal study sample size satisfied statistical power requirements.

Instrument. We used SADCAT (Semi-Automated Dictionary Creation for Analyzing Text) for analysis. Developed by Nicolas et al. [**?**, **?**], SADCAT is a specialized semi-automated English psychological dictionary for studying stereotype content. Constructed based on warmth and competence dimensions from the Stereotype Content Model, it uses WordNet[1] to expand into 16 impression dimension sub-dictionaries including sociability, morality, ability, assertiveness, status, and appearance. Sociability and morality constitute the warmth dimension, while ability and assertiveness constitute the competence dimension. SADCAT may be influenced by seed words[2], WordNet distribution, and semantic generality differences, resulting in unbalanced sub-dictionary lengths ranging from 7 to 2,402 words, though it achieves approximately 82% stereotype coverage.

Operational Definition. Priority effects are typically validated through processing speed, subjective weight, and pragmatic diagnosticity [**?**, **?**]. Processing speed refers to dimensions being processed and recognized earlier in cognition; subjective weight reflects perceivers' tendency to assign greater importance to certain dimensions; pragmatic diagnosticity indicates that dimensions showing higher evaluation consistency across individuals have priority effects. This study adopted processing speed to examine participants' lexical selection tendencies when describing LLMs. Specifically, priority effect was defined as: in participants' LLM descriptions, the dimension with higher coverage was considered more rapidly attended to and processed in cognition, thus representing the prioritized dimension.

Procedure. First, participants were asked via questionnaire to describe their impressions of LLMs using at least three Chinese words, with no restrictions on word type. This yielded 612 descriptive words (including repetitions) from 207 participants. Although some participants used only one or two words, these were retained in the dataset.

Subsequent text processing involved: (1) Translation: converting Chinese words to English using translation tools; (2) Proofreading: review by English professionals to ensure accuracy; (3) Preprocessing: converting to lowercase, removing punctuation, lemmatization[3], and removing stop words[4]; (4) Dimension identification: matching processed words against SADCAT to count words belonging to warmth, competence, sociability, morality, ability, assertiveness, and other dimensions; (5) Coverage calculation: dividing each dimension' s word count by total descriptive words per participant to obtain dimension coverage rates measuring relative prevalence.

We then conducted GLMM analysis using the R package lme4, with coverage as the dependent variable, dimension (e.g., warmth, competence) as fixed effects, participant as random effects, and descriptive word count as weights. Given the unbalanced gender ratio, we initially included gender as a covariate but found no significant effect, so we removed it and refitted the model.

GLMM calculated estimated marginal means for each dimension, providing more accurate effect estimates by assessing average coverage while controlling for other variables. Finally, we compared coverage differences between warmth and competence dimensions to examine their relative impact on LLM perception.

2.1.3 Results We constructed two GLMMs: a two-dimensional model (fixed effects: warmth and competence) and a sub-dimension model (fixed effects: 16 sub-dimensions including sociability, morality, assertiveness, status, ability, beliefs, appearance, etc.; the family dimension was excluded as all participants had zero coverage). Both models showed significant fixed effects (see Table 1). The sub-dimension GLMM had AIC = 2807.34 and BIC = 2904.00, both larger than the two-dimensional model' s AIC = 981.89 and BIC = 993.97. Since AIC and BIC balance model fit and complexity, lower values indicate better fit without excessive complexity, the warmth-competence two-dimensional GLMM demonstrated superior fit.

Next, we calculated estimated marginal means for warmth, competence, and other dimensions, exponentiating them to obtain actual average coverage rates. Results showed highest coverage for warmth and competence dimensions. At the sub-dimension level, the morality facet of warmth and the ability facet of competence showed high coverage, while the sociability facet of warmth and assertiveness facet of competence did not show notably high coverage (see Table 2 and Figure 1 [Figure 1: see original paper]).

Finally, comparing competence and warmth coverage revealed significantly less warmth-related than competence-related content (odds ratio = 2.88, z = 9.51, 95% CI [2.32, 3.59], $p < 0.001$), indicating higher coverage for competence.

2.1.4 Discussion Study 1a used a semi-automated tool to map participant responses through an existing lexical classification dictionary to relevant dimensions. Results supported H1: humans' primary perception dimensions of LLMs align with those of other humans—warmth and competence.

However, unlike human-to-human perception where warmth typically takes priority, this study supported H2, showing competence priority in LLM perception. While this competence-priority effect appears to contradict the warmth-priority effect in the Big Two model, both follow the principle of profit maximization. In human-to-human perception, warmth and competence represent other-profitable and self-profitable traits, respectively, leading observers to prioritize others' warmth to maximize self-interest [?, ?]. In LLM perception, both warmth and competence are viewed as other-profitable traits, with competence' s other-profitable nature being more direct, thus prioritizing competence maximizes self-interest.

An alternative explanation suggests humans perceive LLMs as groups with high outcome dependency. Interpersonal perception research shows that outcome dependency enhances the relevance of others to personal goal achievement, increasing competence' s importance in overall impressions [?, ?]. Outcome dependency refers to the extent to which individuals are affected by others' behaviors and the importance of others' goal achievement to oneself. For example, competence dimensions matter more in impressions of friends than acquaintances [?, ?], more for close friends than distant others [?, ?], and similarly for superiors with dependency relationships [?, ?]. Currently, humans primarily use LLMs for task completion (e.g., seeking advice, information retrieval, solving professional problems) [?, ?], creating high outcome dependency and thus competence priority.

Although Study 1a supported H1 and H2, the SADCAT lexical classification tool may have introduced interference. Developed for human perception, some vocabulary categories may be biased when applied to LLM descriptions. Additionally, as an English dictionary, linguistic and cultural differences may affect results. Therefore, Study 1b will employ traditional trait rating tasks to verify robustness.

2.2 Study 1b: Trait Rating Task

2.2.1 Purpose To avoid SADCAT tool influences, this study directly collected human trait judgments of LLMs through a trait rating task to test H1—that warmth and competence are primary dimensions of LLM perception.

2.2.2 Method **Participants.** We recruited LLM-experienced participants through an online survey platform. Based on statistical power considerations following a 1:5 item-to-participant ratio, we aimed for at least 245 participants, ultimately recruiting 300. After excluding 81 with no LLM experience or who failed to follow instructions, the final sample comprised 219 valid participants (134 males, 61.19%; 85 females, 38.81%). According to MacCallum et al. [?, ?], sample sizes of 100-200 are adequate for factor analysis with high factor loadings and good communalities, while Comrey and Lee [?, ?] consider 200 moderate for factor analysis. Thus, our sample size was acceptable.

Procedure. Based on impression words from Study 1a, we developed an LLM evaluation word pool through: (1) Retaining adjectives while removing nouns, verbs, etc., to focus evaluation on descriptive words and avoid interference; (2) Removing modifying adverbs (e.g., "somewhat," "not very") since participants would later select applicability levels; (3) Removing behavioral descriptions (e.g., "enjoy using") and valence words (e.g., "good") that reflect individual reactions rather than LLM characteristics; (4) Merging antonyms by converting them to positive-valence words when possible, otherwise retaining negative-valence words; (5) Merging synonyms to ensure conciseness.

This process retained 29 words. To avoid omitting relevant LLM characteristics and given recent findings that the Big Five personality model applies to LLMs [?, ?, ?], we incorporated 20 trait words from the Chinese Adjective Big Five Personality Brief Scale [?, ?], which includes two opposing adjectives per item. We randomly selected one word from each item, resulting in a final LLM evaluation word pool of 49 words.

Participants then read instructions for the trait rating task and rated each word' s applicability to LLMs on a 5-point scale (1 = "completely inapplicable," 5 = "completely applicable").

2.2.3 Results Bartlett' s test of sphericity (^{2} = 7926.14, df = 1176, $p < 0.001$) and KMO test (KMO = 0.93) indicated shared latent factors suitable for exploratory factor analysis (EFA). Following Costello and Osborne' s [?, ?] recommendations, we conducted EFA on 49 words using maximum likelihood and varimax rotation in SPSS. Results showed eight factors with eigenvalues > 1, explaining 34.97%, 15.20%, 4.22%, 3.08%, 2.76%, 2.53%, 2.36%, and 2.13% of variance, respectively, cumulatively explaining 67.24%. Based on scree plot criteria [?, ?], we extracted two factors (see Figure 2 [Figure 2: see original paper]).

To simplify factor structure, we iteratively removed: (1) items with low loadings (< 0.50) on all factors; (2) items loading > 0.50 on two or more factors; (3) factors with ≤ 3 items; (4) items with communality < 0.40. This retained 36 words explaining 53.04% of variance. Given the unbalanced gender ratio, we repeated this process for male and female samples separately. Final EFA results are shown in Table 3 .

The two extracted factors corresponded to competence and warmth. Although specific word loadings differed between males and females, both groups clearly showed these two primary factors, with competence explaining more variance (35.62%) than warmth (17.41%).

2.2.4 Discussion Study 1b validated Study 1a results through trait rating tasks. Statistical analyses showed human perception of LLMs concentrates on two dimensions: the first related to intelligent, practical competence words reflecting LLM capability characteristics; the second related to tolerant, sincere warmth words reflecting LLM warmth characteristics. This aligns with Study 1a, confirming warmth and competence as primary LLM perception dimensions.

Notably, some words showed factor loadings diverging from the Big Two model. For example, "patient" —typically a warmth trait in human evaluation representing caring attributes [?, ?]—loaded on the competence dimension in this study. This suggests humans may categorize certain warmth traits as algorithmic performance rather than emotional attributes when evaluating LLMs, possibly interpreting interactive features as programmed design outcomes. Additionally, "cautious" from the Big Five conscientiousness dimension showed negative valence, possibly because the contrasting "bold" was absent, leading participants to interpret "cautious" as "timid" or "hesitant," or because LLMs' need for precise risk-avoidant responses made "cautious" seem like mechanical rather than prudent behavior. Meanwhile, "easy to use" showed negative loading on competence, suggesting high competence may be associated with lower ease of use. Overall, factor analysis revealed that warmth and competence traits attributed to LLMs are not entirely consistent with those attributed to humans.

In summary, although LLM perception primarily revolves around warmth and competence dimensions similar to human perception, details differ. Future research should systematically construct trait word classifications for warmth and competence dimensions specific to LLMs to examine these differences in detail.

---

3 Study 2: Predictive Effects of Warmth and Competence on Attitudes Toward LLMs

3.1 Purpose

This study examines how human perception of LLMs influences attitudinal responses and tests whether dimensional priority effects shift when predicting different attitudes, to verify H3a and H3b.

3.2 Method

3.2.1 Participants To ensure participants could accurately evaluate LLM-related information, we recruited LLM-experienced participants online from the same recruitment batch as Study 1a. After excluding 37 participants without

LLM experience or who failed to follow instructions, the final sample comprised 178 valid participants (105 males, 58.9%; 73 females, 41.1%). All participants volunteered for the study.

To ensure adequate statistical power, we conducted a priori power analysis using G*Power for regression analysis with two predictors and one outcome variable. Setting statistical power at 0.80, effect size f^2 at 0.15 (medium effect), and significance level α at 0.05, we calculated a minimum required sample of 68 participants. Our sample size exceeded this requirement.

3.2.2 Operational Definition This study tested priority effects through subjective weight—the tendency for individuals to assign greater importance to one dimension during evaluation [?, ?]. Specifically, priority effect was defined as: when predicting human attitudes toward LLMs, the dimension with stronger predictive power was considered prioritized.

3.2.3 Instruments and Procedure We used questionnaires to assess continued use intention and liking of LLMs. Continued use intention was measured using three items adapted from Wu and Zhang [?, ?]: (1) "I am willing to use this LLM" ; (2) "I will continue using this LLM in the future" ; (3) "I would recommend this LLM to others." Liking was measured using five items from Bartneck et al.' s [?, ?] Godspeed scale: (1) "This LLM is likeable" ; (2) "This LLM is kind" ; (3) "This LLM is polite" ; (4) "This LLM is pleasant" ; (5) "This LLM is approachable." Both scales used 7-point scoring.

Warmth and competence ratings were collected using Zuo et al.' s [?, ?] 7-point trait scale based on Abele et al. [?, ?], measuring evaluations of both humans and LLMs. The scale includes warmth traits ("warm," "friendly") and competence traits ("intelligent," "competent").

Participants first selected a familiar LLM for evaluation, then completed the scales.

3.3 Results

3.3.1 Common Method Bias Test We used Harman' s single-factor test for common method bias. All measurement indicators were loaded onto a single latent variable for confirmatory factor analysis (CFA). If common method bias existed, all indicators would be highly correlated and the single factor would explain most variance. However, results showed poor fit for the single-factor model ($^2(35) = 479.48$, $p < 0.001$, CFI = 0.69, TLI = 0.60, RMSEA = 0.28, SRMR = 0.17), with CFI and TLI below 0.90, RMSEA above 0.06, and SRMR above 0.08, indicating no common method bias.

3.3.2 Descriptive Statistics Correlation analysis results are shown in Table 4 . All variables—continued use intention, liking, warmth, and competence—showed positive intercorrelations.

3.3.3 Regression Analysis We constructed two linear regression models with continued use intention and liking as dependent variables, and warmth and competence as predictors. Results showed significant predictive effects for both dimensions. Continued use intention and liking had R^2 values of 28.59% and 30.64%, and adjusted R^2 values of 27.77% and 29.85%, respectively (see Table 5 and Figures 3-4 [Figure 3: see original paper][Figure 4: see original paper]). Further analysis revealed that competence had stronger explanatory power than warmth for continued use intention, while warmth had stronger explanatory power than warmth for liking. Interaction terms were non-significant ($p > 0.05$), indicating no warmth × competence interaction effects.

3.4 Discussion

This study supported H3a and H3b, showing that both warmth and competence evaluations positively predicted continued use intention and liking. Competence ($\beta = 0.45$, $p < 0.001$) was a stronger predictor than warmth ($\beta = 0.19$, $p = 0.005$) for continued use intention, while warmth ($\beta = 0.41$, $p < 0.001$) was stronger than competence ($\beta = 0.27$, $p < 0.001$) for liking.

These findings reveal that warmth and competence show different priority effects when predicting distinct attitudes toward LLMs. Competence matters more for functional attitudes, warmth more for affective attitudes. Similar phenomena appear in human perception—people rely more on competence in professional contexts but on warmth in family or social contexts [**?**, **?**]. This suggests attitude type may be a key factor influencing priority effect shifts. However, attitudes are typically mixed functional-affective structures, so directly mapping continued use intention and liking onto functional versus affective attitudes may be oversimplified. Since situational types are more controllable than attitude types, and LLMs are already widely used in functional scenarios (e.g., productivity enhancement) and relational scenarios (e.g., emotional support), future research should examine how situational types influence priority effects.

These results have important implications for LLM product design. Functional interfaces (e.g., professional Q&A systems) should emphasize competence indicators like accuracy and response speed to enhance trust and continued use intention. Affective companion applications (e.g., psychological counseling robots) should strengthen warmth perception through language style and interaction rhythm to enhance emotional dependence and liking.

---

4 Study 3: Comparing Warmth and Competence Ratings of Humans and LLMs

4.1 Purpose

This study compares human ratings of warmth and competence for LLMs versus other humans to test H4a and H4b.

4.2 Method

4.2.1 Participants Participants were from the same recruitment batch as Study 1a (N = 215). After excluding 8 with no LLM experience or who failed to follow instructions, the final sample comprised 207 valid participants (124 males, 59.90%; 83 females, 40.10%). Power analysis using G*Power for paired-samples t-tests ($\alpha = 0.05$, effect size d = 0.50, power = 0.80) indicated a minimum required sample of 34 participants. Our sample (N = 207) exceeded this requirement.

4.2.2 Instruments and Procedure The warmth and competence scale matched Study 2. Participants rated the warmth and competence levels of most people they had encountered and LLMs they had used.

4.3 Results

Paired-samples t-tests showed no significant difference in warmth ratings between LLMs (M = 5.11, SD = 1.23) and humans (M = 5.06, SD = 1.09), t(206) = 0.60, p = 0.551, Cohen' s d = 0.05. However, competence ratings differed significantly, with LLMs (M = 5.16, SD = 1.20) rated higher than humans (M = 4.81, SD = 1.23), t(206) = 3.51, $p < 0.001$, Cohen' s d = 0.29 (see Figure 5 [Figure 5: see original paper] and Table 6).

Given the unbalanced gender ratio and known gender differences in AI attitudes [**?**, **?**], we conducted additional analyses. Male participants rated LLMs higher than humans on both warmth, t(123) = 2.30, p = 0.023, Cohen' s d = 0.21, and competence, t(123) = 2.28, p = 0.024, Cohen' s d = 0.25. Female participants showed no warmth rating difference, t(82) = -1.15, p = 0.255, Cohen' s d = -0.15, but rated LLMs higher on competence, t(82) = 2.71, p = 0.008, Cohen' s d = 0.34 (see Table 6).

4.4 Discussion

Results showed no significant warmth rating difference between LLMs and humans, failing to support H4a. However, LLMs received significantly higher competence ratings than humans, supporting H4b. This reveals that at current technological levels, humans perceive LLMs as superior to humans in competence while matching human average levels in warmth. Comparing this to Durante et al.' s [**?**, **?**] group stereotype research, LLMs in human perception exhibit high competence, moderate warmth—similar to impressions of highly educated groups, elites, wealthy individuals, and men in highly peaceful countries like Denmark.

Existing research supports LLMs' good performance on both dimensions. Although no direct warmth measurement studies exist, research shows LLMs exhibit good agreeableness and empathy [**?**, **?**, **?**], with some studies finding LLMs exceed humans in empathy [**?**, **?**]. Agreeableness relates to trust, openness, helpfulness, humility, and compassion [?, ?], while empathy involves experiencing others' emotions [?, ?], both overlapping with warmth' s friendly, caring characteristics. Competence dimension evidence comes from standardized LLM capability assessments showing outstanding performance [?, ?, ?].

Gender effects warrant attention. Male participants rated LLMs higher than humans on warmth, while females showed no difference—consistent with prior research on gender and AI perception [?, ?]. This may reflect females' greater reliance on emotional feedback in technology use versus males' focus on task completion and utility [?, ?], leading females to be more emotionally invested and sensitive to LLM warmth, while males find warmth "good enough" if functional needs are met.

---

5 General Discussion

This study examines human-machine interaction characteristics from a human perception perspective, against the backdrop of new-generation AI agents like LLMs becoming deeply integrated into human society. By constructing a theoretical framework for human perception of LLMs, it provides support for theoretical construction and empirical research on human-AI relationships in the era of human-machine symbiosis.

Grounded in the Big Two model of social cognition content, three studies examined warmth and competence dimensions in LLM perception. Study 1 found warmth and competence as primary LLM perception dimensions, with competence prioritized over warmth in general contexts. Study 2 revealed that both dimensions positively predicted continued use intention and liking, with competence more predictive of continued use intention and warmth more predictive of liking. Study 3 showed no significant warmth rating difference between LLMs and humans, but significantly higher competence ratings for LLMs. Overall, the human perception framework for LLMs resembles that for other humans, centering on warmth and competence, but with competence priority in general contexts. In human perception, LLMs appear as high-competence, moderate-warmth entities rather than matching general human levels.

5.1 Theoretical Contributions

This study makes two core theoretical contributions: extending the Big Two model' s applicability and broadening theoretical perspectives in human-computer interaction.

First, it establishes a preliminary theoretical framework for human perception of LLMs, expanding the Big Two model from humans to LLMs. Regarding dimensions, Fiske et al. [?, ?] explained warmth and competence as primary stereotype content dimensions from an evolutionary perspective, arguing they correspond to intent detection and threat capability detection, respectively, thus serving survival [?, ?]. DPM functionally explains these dimensions as serving perceivers' goals [?, ?], which are more specific than evolutionary survival goals. Regardless of explanation, as LLM technology advances rapidly and may pose potential threats while affecting goal achievement, warmth and competence remain primary LLM perception dimensions. This study also explored priority effects, finding competence priority rather than warmth priority in general LLM perception contexts—while still following the profit maximization principle. This principle states that warmth, as an other-profitable trait, is prioritized in human perception to maximize self-interest [?, ?]. In LLM perception, both dimensions are other-profitable, but competence' s other-profitability is more direct, leading to competence priority. However, priority effects shift in attitude prediction: competence priority for functional attitudes (continued use intention) and warmth priority for affective attitudes (liking). Similar shifts occur in human perception across different contexts and targets [?, ?, ?, ?].

Second, this study broadens human-computer interaction theory, particularly expanding the theoretical foundation for user experience. Previous AI user experience research relied primarily on Technology Acceptance Model (TAM) and Unified Theory of Acceptance and Use of Technology (UTAUT), focusing on performance expectancy, ease of use, and traditional factors [?, ?, ?, ?]. Recent research has expanded variables affecting human-AI attitudes to include contextual, user characteristic, and machine characteristic factors—for example, showing that trust tendencies, robot credibility features, and service contexts affect interaction trust with social service robots [?, ?]. However, this expanding variable set has increased model complexity without forming a more concise, systematic explanatory framework. This study introduces warmth and competence—two core social cognition dimensions—into human-LLM interaction research, systematically examining their impact on user attitudes and providing a new theoretical perspective for optimizing human-machine interaction and enhancing system controllability and safety.

5.2 Practical Implications

These theoretical contributions yield practical significance in both social and technological domains.

Socially, incorporating LLMs into the Big Two framework provides theoretical foundations for designing more reasonable institutional arrangements and ethical norms for the human-machine symbiosis era. Institutionally, given LLMs' strong warmth and competence performance, policymakers must consider integrating LLMs into existing legal systems—for example, developing specialized LLM legislation clarifying legal status, rights and obligations, and regulatory mechanisms to ensure LLM development serves public interests. As LLM capabilities impact labor markets, this study encourages designing employment policies from a "human-LLM symbiosis" perspective, such as providing skills training and transition assistance to reduce technological inequality. Ethically, the study reminds developers to consider LLMs'ethical and social responsibilities alongside technological progress, designing behavior patterns and interaction methods that avoid negative social impacts and promote human welfare.

Technologically, this study provides new theoretical foundations for LLM design, development, and application. Regarding explainability and controllability, the theoretical model reveals warmth and competence' s importance in LLM perception, offering a new framework for understanding and controlling LLM behavior. This helps developers design more interpretable and controllable algorithms, reducing unpredictability and enhancing safety and reliability. It also guides technical optimization—developers can focus on improving LLM performance on these dimensions, such as enhancing dialogue fluency to boost competence perception or designing more empathetic responses to strengthen warmth perception, thereby increasing user acceptance and trust. Additionally, developers can design LLMs with specific social roles (e.g., educational assistants, psychological companions) to better serve human society.

5.3 Limitations and Future Directions

First, expanding from dimensions to facets. Recent social cognition research increasingly examines warmth and competence facets to explain theoretical differences and explore psychological mechanisms [**?**, **?**]. DPM divides warmth into morality and friendliness facets, competence into ability and assertiveness facets [**?**, **?**]. Study 1a found LLM warmth perception concentrated on morality and competence perception on ability, while Study 1b showed less clear facet differences, possibly due to fewer evaluation words. The classification of "patient" as competence in Study 1b also suggests LLM competence dimensions may encompass unique extensions beyond human perception. Future research should use web scraping to extract richer, more authentic descriptive vocabulary from natural contexts to expand word pools and explore facet applicability in LLM perception, refining theoretical frameworks and dimensional boundaries.

Second, contextualized and dynamic analysis of dimensional priority effects. We found priority effects shift across attitude predictions. As LLMs are widely used in both efficiency-enhancing and emotional support contexts, future research must examine how situational types affect priority effects. Experimental studies should explore mechanisms and conditions for these shifts, with key factors potentially including situational type (functional vs. relational contexts).

Third, exploring complex mechanisms influencing LLM perception. This study did not deeply investigate cues (causes) and mechanisms affecting LLM perception. Future research should refine these aspects to construct a complete theoretical framework.

Finally, this study prompts societal re-examination of the "human" category. While limited to examining LLMs' image in human perception and comparing it to human-to-human perception, without addressing LLMs' ontological human-likeness, both the similarity in warmth/competence perception and LLMs' evolving human-likeness inevitably change their social roles in human-machine symbiosis society, potentially blurring human-machine boundaries. As LLM functions improve, developers face dilemmas: imbuing them with human-like traits while ethically restricting them to tool status. Yet humans may still perceive LLMs as quasi-human in specific interactions. The deeper question is whether the concept of "human" is sociologically changing as technology advances. If LLMs are perceived as possessing human mental capabilities, has the definition of "human" been generalized? Recent research suggests that leveraging human perceptual identity can extend ontological parity to non-human elements, indicating the malleability of the "human" concept in social cognition [**?**, **?**]. Future research could use Mind Perception Theory to investigate LLMs' ontological human-likeness from experiential (emotional/perceptual capacity) and agency (autonomous action/self-control) dimensions [**?**, **?**, **?**]. Though overlapping partially with the Big Two model [**?**, **?**, **?**, **?**], Mind Perception Theory emphasizes the cognitive construction of "mind." Entities lacking either dimension are not granted "mind" —a traditionally human-exclusive feature. Investigating LLMs' ontological human-likeness and its social interaction manifestations may witness profound evolution of the "human"category and advance theoretical construction of human-machine symbiosis society.

6 Conclusion

This study' s main conclusions are: First, warmth and competence are primary dimensions of human LLM perception, with competence prioritized over warmth in general contexts. Second, both warmth and competence ratings positively predict continued use intention and liking, with competence more predictive of continued use intention and warmth more predictive of liking. Third, human warmth ratings for LLMs do not differ significantly from those for other humans, but competence ratings for LLMs are significantly higher than for humans.

References

Abele, A. E., & Brack, S. (2013). Preference for other persons'traits is dependent on the kind of social relationship. *Social Psychology, 44*(2), 84–94.

Abele, A. E., Bruckmüller, S., & Wojciszke, B. (2014). You are so kind -and I am kind and smart: Actor -Observer Differences in the Interpretation of On-going Behavior. *Polish Psychological Bulletin; 2014; No 4, 45*(4), 457–470.

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the Social World: Toward an Integrated Framework for Evaluating Self, Individuals, and Groups. *Psychological Review, 128*(2), 290–314.

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the Fundamental Content Dimensions: Agency with Competence and Assertiveness-Communion with Warmth and Morality. *Frontiers in Psychology, 7*(1). Article 1810.

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*(5), 751-763.

Abele, A. E., & Wojciszke, B. (2014). Communal and Agentic Content in Social Cognition: A Dual Perspective Model. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol 50, pp. 195-255). Elsevier Academic Press Inc.

Abele, A. E., & Wojciszke, B. (Eds). (2018). *Agency and communion in social psychology.* Routledge.

Bai, X., Ramos, M. R., & Fiske, S. T. (2020). As diversity increases, people paradoxically perceive social groups as more similar. *Proceedings of the National Academy of Sciences, 117*(23), 12741-12749.

Bartneck, C., Kulic, D., Croft, E., & Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics, 1*(1), 71-81.

Bodroža, B., Dinić, B. M., & Bojić, L. (2024). Personality testing of large language models: limited temporal stability, but highlighted prosociality. *Royal Society Open Science, 11*(10), 240180.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ···Amodei, D. (2020). Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (pp. 1877-1901). Curran Associates Inc.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ···Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv.* https://doi.org/10.48550/arXiv.2303.12712

Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017, March). The Robotic Social Attributes Scale (RoSAS): Development and Validation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 254-262). ACM.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276.

Chi, O. H., Jia, S., Li, Y., & Gursoy, D. (2021). Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. *Computers in Human Behavior, 118*, 106700.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis, 2nd ed* (pp. xii, 430). Lawrence Erlbaum Associates, Inc.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92*(4), 631-648.

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J. P., ···Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology, 48*(1), 1-33.

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A Review of the Concept. *Emotion Review, 8*(2), 144-153.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Mis Quarterly, 13*(3), 319-340.

de Ruyter, B., Saini, P., Markopoulos, P., & van Breemen, A. (2005). Assessing the effects of building social intelligence in a robotic interface for the home. *Interacting with Computers, 17*(5), 522-541.

dos Santos, R. P. (2023). Enhancing Physics Learning with ChatGPT, Bing Chat, and Bard as Agents-to-Think-With: A Comparative Case Study. *arXiv*. https://doi.org/10.48550/arXiv.2306.00724

Durante, F., Fiske, S. T., Gelfand, M. J., Crippa, F., Suttora, C., Stillwell, A., ···Carlsson, R. (2017). Ambivalent stereotypes link to peace, conflict, and inequality across 38 nations. *Proceedings of the National Academy of Sciences, 114*(4), 669-674.

Fiske, S. T. (1992). Thinking is for doing: portraits of social cognition from daguerreotype to laserphoto. *Journal of Personality and Social Psychology, 63*(6), 877-889.

Fiske, S. T. (2018). Stereotype Content: Warmth and Competence Endure. *Current Directions in Psychological Science, 27*(2), 67-73.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77-83.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878-902.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619-619.

Gray, K., & Wegner, D. M. (2010). Blaming God for Our Pain: Human Suffering and the Divine Mind. *Personality and Social Psychology Review, 14*(1), 7-16.

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and*

Evolution, 7(4), 493-498.

Heaven, P. C. L., Ciarrochi, J., Leeson, P., & Barkus, E. (2013). Agreeableness, conscientiousness, and psychoticism: Distinctive influences of three personality dimensions in adolescence. *British Journal of Psychology, 104*(4), 481-494.

Heflick, N. A., Goldenberg, J. L., Cooper, D. P., & Puvia, E. (2011). From women to objects: Appearance focus, target gender, and perceptions of warmth, morality and competence. *Journal of Experimental Social Psychology, 47*(3), 572-581.

Huang, J., Wang, W., Lam, M. H., Li, E. J., Jiao, W., & Lyu, M. R. (2023). Revisiting the Reliability of Psychological Scales on Large Language Models. *arXiv*. https://doi.org/10.48550/arXiv.2305.19926

Jiang, G. Y., Xu, M. J., Zhu, S. C., Han, W. J., Zhang, C., & Zhu, Y. X. (2023, December). Evaluating and Inducing Personality in Pre-trained Language Models. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*. Neural Information Processing Systems (Nips).

Kosinski, M. (2024). Evaluating large language models in theory of mind tasks. *Proceedings of the National Academy of Sciences, 121*(45), e2405460121.

Lee, Y. K., Suh, J., Zhan, H., Jessy Li, J., & Ong, D. C. (2024). Large Language Models Produce Responses Perceived to be Empathic. *arXiv*. https://doi.org/10.48550/arXiv.2403.18148

Luo, J., & Dai, X. (2018). Chinese Adjective Big Five Personality Scale Brief Version Development. *Chinese Journal of Clinical Psychology, 26*(04), 721-726.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84-99.

McKee, K. R., Bai, X. C. Z., & Fiske, S. T. (2023). Humans perceive warmth and competence in artificial intelligence. *Iscience, 26*(8), 27.

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI, 4*(1), Article 21.

Mohammad, Y., & Nishida, T. (2015). Why Should We Imitate Robots? Effect of Back Imitation on Judgment of Imitative Skill. *International Journal of Social Robotics, 7*(4), 497-512.

Mori, M. (2012). The Uncanny Valley. *Ieee Robotics & Automation Magazine, 19*(2), 98-100.

Nicolas, G., Bai, X. C. Z., & Fiske, S. T. (2021). Comprehensive stereotype content dictionaries using a semi-automated method. *European Journal of Social Psychology, 51*(1), 178-196.

Nomura, T., Kanda, T., & Suzuki, T. (2006). Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. *Ai*

& Society, 20(2), 138–150.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ···Ray, A. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35*(27730–27744), Article 2011.

Scheunemann, M. M., Cuijpers, R. H., & Salge, C. (2020). Warmth and Competence to Predict Human Preference of Robot Behavior in Physical Human-Robot Interaction. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 1340–1347). 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN).

Sorin, V., Brin, D., Barash, Y., Konen, E., Charney, A., Nadkarni, G., & Klang, E. (2024). Large Language Models and Empathy: Systematic Review. *Journal of Medical Internet Research, 26*(10), Article e52597.

Strachan, J. W. A., Albergo, D., Borghini, G., Pansardi, O., Scaliti, E., Gupta, S., ···Becchio, C. (2024). Testing theory of mind in large language models and humans. *Nature Human Behaviour, 8*(7), 1285–1295.

Sundararajan, L., Wu, M. S., Ho, W.-T., Sun, J. C., Leung, C. P., & Rich, G. J. (2022). Expanding our kind: A pan-cultural study of the animistic principle of ontological parity. *The Humanistic Psychologist, 50*(3), 476–494.

Tae, M., & Lee, J. (2020). The Effect of Robot’ s Ice-breaking Humor on Likeability and Future Contact Intentions. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 462–464). ACM.

Uchronski, M. (2008). Agency and communion in spontaneous self-descriptions: Occurrence and situational malleability. *European Journal of Social Psychology, 38*(7), 1093–1102.

Venkatesh, V., & Morris, M. (2000). Why Don’ t Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior. *Mis Quarterly, 24*(1), 115–139.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *Mis Quarterly, 27*(3), 425–478.

Wang, J., Ma, W., Sun, P., Zhang, M., & Nie, J.-Y. (2024). Understanding User Experience in Large Language Model Interactions. *arXiv*. https://doi.org/10.48550/arXiv.2401.08329

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who Sees Human?: The Stability and Importance of Individual Differences in Anthropomorphism. *Perspectives on Psychological Science, 5*(3), 219–232.

Waytz, A., & Young, L. (2014). Two motivations for two dimensions of mind. *Journal of Experimental Social Psychology, 55*, 278–283.

Webb, T., Holyoak, K. J., & Lu, H. J. (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour, 7*(9), 1526–1541.

Wojciszke, B., & Abele, A. E. (2008). The primacy of communion over agency and its reversals in evaluations. *European Journal of Social Psychology, 38*(7), 1139–1147.

Zhao, Y. K., Huang, Z., Seligman, M., & Peng, K. P. (2024). Risk and prosocial behavioural cues elicit human-like response patterns from AI chatbots. *Scientific Reports, 14*(1), Article 7095.

佐斌, 代涛涛, 温芳芳. (2015). 社会认知内容的”大二”模型. 心理科学*, 38*(04), 1019–1023.

佐斌, 温芳芳, 吴漾. (2018). 群际评价中热情与能力关系的情境演变: 评价意图与结果的作用. 心理学报*, 50*(10), 1180–1196.

武晓宇, 张大伟. (2023). 学术用户移动视觉搜索的持续使用意愿研究——基于 UTAUT2-ECT 整合模型的实证分析. 情报杂志*, 42*(08), 164–176.

罗杰, 戴晓阳. (2018). 中文形容词大五人格量表的初步编制 : 简式版的研制. 中国临床心理学杂志*, 26*(04), 721–726.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.