Anthropomorphic Cues and Emotional Design in Social AI: Introducing and Validating the Cue-Perception-Relation (CPR) Model

Author

Ruolin Zhao

10946625 | 242116

Supervisor

Marco Ajovalasit

Course

Msc. Digital and Interaction Design

POLITECNICO
MILANO 1863

Truth is always an inexplicable inner contact.
— Clarice Lispector, The Hour of the Star

Abstract (EN)

Anthropomorphic cues are widely employed in social AI to enhance emotional engagement, yet their relational implications remain structurally underexplored. This study conceptualizes anthropomorphic design as a layered relational mechanism and proposes the Cue–Perception–Relation (CPR) model, linking anthropomorphic cues (appearance, behavior & expression, identity) to users' interpretive perceptions and anticipated human–AI relational outcomes. Through complementary designer-oriented and user-oriented scenario-based experiments, the model is empirically validated using regression analyses, demonstrating significant cross-layer relationships and perceptual mediation effects. Building on this calibration, a refined set of CPR-aligned principles is derived to guide the context-sensitive application of anthropomorphic cues. By situating anthropomorphic design at the intersection of social AI and emotional design, this research advances a structurally grounded and empirically supported model for shaping user perception, relational expectations, and boundary management in human–AI interaction.

Keywords: Anthropomorphism; Social AI; Emotional Design; Human–Computer Interaction; Human–AI Relationship; Context-Sensitive Design; User Experience; Empirical Design Validation

Abstract (IT)

I segnali antropomorfici sono ampiamente impiegati nei sistemi di social AI per favorire il coinvolgimento emotivo; tuttavia, le loro implicazioni percettive e relazionali rimangono ancora strutturalmente poco esplorate. Il presente studio concettualizza il design antropomorfico come un meccanismo stratificato e propone il modello Cue–Perception–Relation (CPR), che collega i segnali antropomorfici (aspetto, comportamento ed espressione, identità) alle percezioni interpretative degli utenti e agli esiti relazionali anticipati nell'interazione uomo–AI. Attraverso esperimenti complementari basati su scenari, orientati sia ai designer sia agli utenti, il modello viene validato empiricamente mediante analisi di regressione, evidenziando relazioni significative tra i diversi livelli del framework. A partire da tale calibrazione, viene definito un insieme raffinato di principi allineati al CPR per guidare l'applicazione contestuale dei segnali antropomorfici. Collocando il design antropomorfico all'intersezione tra social AI ed emotional design, questa ricerca propone un quadro strutturalmente fondato e supportato empiricamente per modellare la percezione dell'utente, le aspettative relazionali e la gestione dei confini nell'interazione uomo–AI.

Parole chiave: Antropomorfismo; Social AI; Emotional Design; Interazione Uomo–Computer; Relazione Uomo–AI; Design Sensibile al Contesto; Esperienza Utente; Validazione Empirica del Design

Index

- List of Tables
- List of Graphs
- List of Figures
- Scenario-related Parameters
- Design-level Parameters
- Principle-level Parameters
- Design Prediction Parameters
- User CPR measures - Perception
- User CPR measures - Relation
- Consent Form - Study I
- Consent Form - Study II
- Design Systems
- Likert - Study I
- Likert - Study II
- Semi-structured Interview - Study I
- Semi-structured Interview - Study II

Acknowledgements

I would like to express my gratitude to Professor Marco Ajovalasit for his role as the supervisor of this thesis. My time at Politecnico di Milano has been defined by the friends I met along the way. Thank you for the intellectual companionship and the support that allowed me to grow both personally and academically.

I am also deeply grateful to my parents; though we are far apart, your unwavering love remains the foundation of my life. I also wish to thank myself for maintaining a genuine passion for this topic and for the dedication required to complete this work.

Finally, this thesis is deeply rooted in my continuous reflection on design practice and observations of the world around us. I believe that the essence of design lies in the commitment to "keep asking why" while engaging in hands-on practice. By maintaining an open mind and embracing the cross-pollination between design and other disciplines, we can generate enduring value, especially in this era of rapid technological innovation and constant change.

1. Introduction

Over the last several decades, artificial intelligence has progressively integrated into society, evolving from theoretical concepts into interactive systems embedded within human experiences. A central aspect of this evolution has been the integration of human-like traits in AI, which has been a persistent focus since its early conceptualizations. The work of Alan Newell and Herbert A. Simon in 1972 laid the foundation for early AI research, conceptualizing human cognition as an information processing system and suggesting the potential for replicating such processes in machines (Arthur I. Miller, 2019). The influence of anthropomorphism in AI development is prominent not only in technological advancements but also in its portrayal across media, shaping societal perceptions and user acceptance. In particular, the design of social AI products leverages anthropomorphic traits to foster emotional trust and improve user engagement (Muriel Reuter, 2024).

Despite the extensive research on the role of anthropomorphism in social AI, a gap remains between theoretical frameworks and practical design applications. Much of the existing work has focused on the classification and impact of anthropomorphic traits in various AI applications, such as chatbots and robots, emphasizing user cognition, trust, and satisfaction (Goa & Sundar, 2019; Darling, 2017; Xie, 2023; Alabed et al., 2022). However, the inconsistent and sometimes ambiguous application of anthropomorphic cues in design often leads to user confusion and unmet expectations. Therefore, understanding anthropomorphic cues as a "language" in design, rather than simply a set of tools for problem-solving, becomes essential in fostering meaningful emotional connections. This shift aligns with the Design for Meaning (DfM) approach, which reinterprets human-centred design by viewing people as a valuable source of inspiration for new meanings (Ajovalasit and Giacomin, 2025). By focusing on the "reasons why" people value their daily activities, DfM provides a lens to rethink about anthropomorphic design in social AI with fundamental questions: "Why do we want to do it?" and "Do others want it, too?"

This paper redefines anthropomorphism in social AI as a relational mechanism that facilitates these meaning-driven interactions, going beyond superficial design features. By introducing a structured Cue-Perception-Relation (CPR) model, the research provides a framework for learning about people's evoked meaning patterns and understanding the connections between design cues, user perceptions and relational outcomes. Empirical evaluations through both designer- and user-centered experiments validate and refine this model, ultimately offering design principles for the calibrated and context-sensitive application of anthropomorphic cues in social AI. Situated within both theoretical inquiry and design practice, this work bridges the gap between academic debates and the practical challenges in social AI development. The following research agenda explores key questions spanning across AI, anthropomorphism, social AI, emotional design, and meaning-driven design, providing the intellectual foundation and scope of this study. The following Table 1.1 shows further details.

Construct	Research Questions Ladder
Artificial Intelligence	• How does artificial intelligence fundamentally differ from human intelligence in terms of agency, cognition, and responsibility? • What dominant misconceptions shape public understanding of AI and its societal implications? • How does the persistent analogy between neuroscience and AI influence expectations about machine intelligence and behavior?
Anthropormophism	• How is anthropomorphism defined across disciplines, and what are its psychological and cultural origins? • In which domains and contexts is anthropomorphism most likely to emerge in human–AI interaction? • How can anthropomorphic features in AI be systematically categorized? • What functional and experiential goals does anthropomorphic design serve in AI systems? • How do anthropomorphic cues shape users' perception, emotional response, and interpretation of AI capabilities? • What positive and negative consequences does anthropomorphic design produce at individual, social, and relational levels?
Social AI	• What distinguishes social AI systems from other intelligent or interactive technologies? • How can the social qualities of AI systems be meaningfully evaluated beyond task performance and usability? • How can social AI be designed to participate in social interactions without encouraging misleading interpretations of human-like sociality?
Emotional Design	• How does emotional design operate within AI systems, and how does it differ from emotional design in traditional interactive products? • In what ways do anthropomorphic features contribute to emotional engagement between users and AI systems? • Under what conditions does anthropomorphic emotional design facilitate emotional connection, and when does it lead

	to emotional misrecognition or transferential engagement?
Design for Meaning	• What motivates designers to incorporate anthropomorphic cues as a means of eliciting emotional responses in AI systems? • How do designers conceptualize the relationship between anthropomorphism, emotional connection, and user interpretation? • What factors shape users' emotional attachment to anthropomorphic AI systems? • How can emotional connection be distinguished from empathic transference and emotional dependency in users' behavioral and relational responses?

Table 1.1: Research agenda

2. Rethinking Anthropomorphism in Social-AI Products

2.1 Anthropomorphism and Its Psychological Motivations

Anthropomorphism long predates the machine age. It is recognized as a complex phenomenon applied not only to machines but also to animals, plants, natural landscapes, and tools. The term "anthropomorphism" derives from the Greek "anthropos" (meaning human) and "morph" (meaning form), signifying the projection of human characteristics onto non-human entities. (Epley, 2007)

Attributing anthropomorphic qualities to objects functions as a cognitive strategy through which people make sense of non-human entities within human social contexts, projecting emotional, cognitive, and other agent-like properties onto them. This inductive reasoning process is driven by innate psychological motivations, specifically the needs to experience competence and social connection. By integrating contextual cues with social knowledge, individuals establish meaningful relationships with non-human entities.

In modern society, people frequently interact with non-human entities in daily life, primarily driven by social connection, motivational influences, and perceived similarities. MacInnis and Folkes (2017) explored anthropomorphism in brands, noting consumers rebuild meaning and social stability through brand anthropomorphism. Diane and Virve (2018), in their analysis of the children's film Toy Story, found that animated anthropomorphic cues, such as nicknames and dynamic movements, help children enhance their understanding of referential expressions. Suan Yuan and Peng Yu's robotic artwork

Can’t Help Myself (2016) mimics human movements as it continuously sweeps up a deep red liquid seeping across the floor.With over 50 million views on TikTok, the piece elicited strong empathy from viewers. Comments like "It looks so tired" project the human emotion of fatigue onto this machine performing meaningless labor, making the non-human entity relatable.

2.2 Artificial Intelligence and Anthropomorphism

2.2.1 Artificial Intelligence and Human Intelligence

The pervasive analogy between artificial and human intelligence has significantly influenced AI's development. Popular culture reinforces this perspective, shaping public perception through anthropomorphic narratives, (Bartneck, 2013) such as the emulated self-awareness in the movie Her or the emotional androids of the game Detroit: Become Human. Within the field, the “computer-brain” analogy and the human-centric Turing Test have historically defined intelligence as the capacity to mimic human behavior. However, this anthropocentric focus potentially restricts a deeper understanding of AI’s unique, non-human characteristics.

As technology evolves, AI's role has transcended its origins as a mere tool, functioning instead as a substitute worker, collaborator, or curator. (Miller, 2019) This emergence of diverse identities triggers anxieties regarding labor displacement and the erosion of human uniqueness. Such tensions necessitate a critical re-evaluation of the conventional human-AI analogy and the intent behind anthropomorphic design, prompting a fundamental question: Is the purpose of AI to emulate or to replace humanity?

Returning to the Turing Test’s "imitation game," Bratton (2015) critiques its inherent anthropocentric bias while acknowledging the strategic value of mimicry. For Bratton, the significance of imitation lies not in achieving superficial similarity, but in enabling machines to transcend their mechanical nature to gain a distinct "difference and vitality". Shneiderman (2022) warns that conflating humans and computers risks diminishing the complexity and creativity unique to the human experience. Thus, AI anthropomorphism is more than mere mimicry; it acts as a bridge between human and machine, possessing a generative significance that responds to human projections while dismantling the binary opposition between the two.

2.2.2 The Role of Anthropomorphic Design in AI

Anthropomorphic design has generated a spectrum of scholarly responses, ranging from enthusiasm to skepticism regarding its cognitive and emotional impact. Brian (2003) posits that the motivation for anthropomorphism in machines should be to enrich human social spaces rather than to engineer synthetic humans. Physically, anthropomorphic configurations, such as the bipedal mechanisms of Boston Dynamics' Atlas, allow machines to navigate human-centric environments like stairs and doorways with greater efficiency. Socially, the judicious application of anthropomorphic cues serves as a catalyst for human-machine socialization, prioritizing interactional quality over mere aesthetic form.

Some scholars emphasize that aligning AI systems with existing human schemas enhances intuitiveness and reduces the cognitive load required for interaction. For instance, human-like perspectives in gaming environments have been shown to promote prosocial cooperation, while anthropomorphic expressions and gestures in service robots help users rapidly develop accurate mental models, increasing familiarity with system operations.

P1: Anthropomorphic cues that help users establish mental models of systems more rapidly and coherently, thereby enhancing cognitive adaptability.

Furthermore, anthropomorphism fosters emotional bonds, intimacy, and trust between human and AI. Research by the Nielsen Norman Group (2023) indicates that anthropomorphic design can lead users to perceive AI systems as more effective, resulting in more enjoyable interactions. This is supported by comparative studies on in-car assistants, where participants exhibited significantly higher levels of trust in the safety of autonomous vehicles equipped with anthropomorphic voices. These phenomena are grounded in the theoretical frameworks of the Media Equation (1996), which suggest that anthropomorphic features trigger innate empathy mechanisms. By perceiving artificial systems as social partners, users infuse the interaction with warmth and sociality—a response that is particularly vital for establishing sustained emotional connections and long-term trust.

P2: Anthropomorphic expression stimulate human's innate social responses, promoting emotional connection and trust-building during interactions.

2.2.3 Framing Anthropomorphic Design in AI: Addressing Concerns and Risks

Research indicates a natural human tendency to anthropomorphize technology regardless of its design. (Sung et el, 2007) In response, Kate (2017) proposes framing as a solution: designers and developers can influence human responses to machines through deliberate, purposeful anthropomorphic design. Addressing the resulting concerns requires following aspect.

a. Application Domains of Anthropomorphic Design

While the anthropomorphic robot market is projected for massive growth, these traits remain a small subset of the broader AI field. Although such traits can boost adoption and interaction, misguided applications, like the early "Tillie the Teller" ATM deployed to enhance customer trust, banks featured a smiling female illustration on these machines, accompanied by voice prompts, aiming to reduce user discomfort with machine interactions through anthropomorphic cues. However, this design failed to achieve its intended effect, instead provoking user aversion and distrust.

Furthermore, research suggests that individuals often demonstrate a higher degree of leniency toward errors committed by anthropomorphic robots (Yam, 2020). This phenomenon is primarily driven by the tendency to perceive these machines as "social others". However, when robots are employed in high-precision or high-stakes environments, the inclusion of social markers such as names or gender may obscure their functional identity as tools, potentially leading users to miscalibrate their assessment of the robot's technical capabilities.

Shneiderman (2022) argues that, in contrast to anthropomorphic AI systems, mainstream AI remains characterized by consistency, predictability, and controllability. These attributes facilitate user mastery, thereby enhancing both operator confidence and a sense of personal accountability. While AI systems equipped with social-emotional intelligence may flourish in entertainment or interpersonal contexts, they frequently become a source of distraction or irritation in task-oriented applications.

Thus, rather than serving as a universal design solution, anthropomorphism functions best as a contextual amplifier—enhancing engagement, empathy, or approachability in domains such as social robots, entertainment systems, and certain service-oriented applications.

P3: Anthropomorphic design should be applied with contextual sensitivity, primarily enhancing emotional and social interaction.

b. Manipulation Risks: Privacy, Dependency, Social Orientation and Bias

Finch: Self-care cat is an application that encourages users to develop self-care habits by feeding an electronic pet. During habit formation, the pet gradually grows, actively feedbacks and establishes an emotional connection with the user through a positive partnership.While using anthropomorphic design to emotionally motivate people toward positive habits seems beneficial, what else might it drive users to do? Purchase products? Renew subscriptions? Share personal secrets?

Although anthropomorphic design can encourage goal achievement, but it also introduces risks of manipulation.

- Privacy: Emotional attachment may incentivize users, particularly vulnerable groups, to disclose personal data in exchange for functional feedback. (Puntoni et al, 2021)
- Dependency: Long-term relationship building with anthropomorphic agents may foster over-reliance on AI, which can erode individual agency and cognitive autonomy, a phenomenon sometimes termed "digital dementia".
- Social Orientation: Anthropomorphic design should facilitate, not replace authentic interpersonal interactions. A key example is the *ElliQ robot*. Rather than acting as a permanent replacement for friends or family, it functions as a "social mediator" that encourages users to maintain their existing social bonds and promotes active social engagement.
- Bias: Anthropomorphic design may entrench broader societal prejudices by coding them into daily human-AI interactions. The proliferation of feminized AI (e.g., *IRON*) often reinforces harmful gender stereotypes associating women with service and emotional labor. Designers have the opportunity for “reverse innovation” by adopting gender-neutral or diverse anthropomorphic expressions.

Overall, while anthropomorphic AI effectively promotes social well-being and connectivity, its human-like interface introduces critical risks of dependency and the reinforcement of systemic biases. Crucially, the same traits that enhance accessibility can be weaponized for opaque manipulation, subtly eroding user autonomy and the capacity for independent judgment.

P4: Anthropomorphic AI should function as a social facilitator that encourages real-world engagement, thereby safeguarding user agency and preventing artificial dependency.

c. Human-Machine Boundaries and User Expectations

Concerns about users confusing AI for humans appear premature, as anthropomorphism is typically a contextual explanation rather than an identity equivalence. (Epley, 2007) Concurrently, research indicates that explicitly presenting a system's AI identity or capability boundaries within an interface can significantly reduce the likelihood of human-machine confusion. (Luger & Sellen, 2016) The more urgent challenge is the gap between user expectations and the actual functional capabilities of the AI.

As Foner (1997) noted, highly anthropomorphic designs often substantially elevate users' expectations of intelligent systems, leading them to unconsciously evaluate system performance against human capability standards; when systemic errors occur, users may feel a sense of "betrayal," amplifying distrust. However, as previously discussed, fundamental differences exist between artificial and human intelligence. Successful anthropomorphic design should avoid misleading users into believing AI possesses human-like intentionality and comprehension, instead maintaining the visibility of its artificial attributes.

To balance capabilities with expectations, Brian's (2003) anthropomorphic design principles are distilled into three primary design orientations.

- Restrained Appearance: Ensure form serves function and avoid the "uncanny valley" by retaining subtle artificial cues .
- Behavioral Paradigms: Focus on human-like linguistic rhythms and conversational logic rather than just visual avatars.
- Distinct Identity: Position AI as a "unique artificial entity" with clear role definitions rather than an incomplete human substitute .

P5: Anthropomorphic AI design should regulate user expectations regarding system capabilities and roles through restrained visual presentation, comprehensible behavioral patterns, and a clear artificial identity positioning.

In conclusion, the discourse should shift from generalized anxiety toward systematic research that establishes verifiable design specifications, ensuring AI delivers value while maintaining controllable risks.

The above discussion reveals that anthropomorphic design in AI yet inevitably carries multiple concerns such as privacy risks, dependency and expectation imbalance. Rather than rejecting or avoiding anthropomorphism with a blanket approach, it is preferable to apply it human-centeredly based on a thorough understanding of its mechanisms and boundary conditions. In conclusion, the discourse should shift from generalized anxiety toward systematic research that establishes verifiable design specifications, ensuring the design delivers value while maintaining controllable risks.

2.2.4 Classification of Anthropomorphic Cues in AI

Systematic classification of anthropomorphic cues is a prerequisite for the standardization and operationalization of design. By categorizing these cues, designers can clarify their underlying mechanisms and determine appropriate applications within user perception. Existing literature offers diverse criteria: some distinguish between physical and psychological dimensions, while others, such as the *MAIN model*, focus on information presentation. Building on Go and Sundar's (2019) framework of visual, identity, and communication cues, this study synthesizes a tripartite classification: Appearance, Behavior & Expression, and Identity cues. (Table 2.1)

Appearance Static & Morphological	**Behavior&Expression** Dynamic & Interactive	**Identity** Sociological
• Body/Avatar • Face • Style	• Movement • Facial Expression • Voice & Tone • Conversation • Emotion	• Name • Gender • Backstory • Role • Personality
To shape the user's initial impression of humanness by providing visual indicators.	To convey dynamic social qualities through motion, interactive responsiveness and emotional expressivity.	To establish the agent's social role and guiding user's expectation about who the agent is.

Table 2.1: Classification of Anthropomorphic Cues in AI

1. Appearance Cues

Appearance cues comprise all visually perceptible features, including humanoid silhouettes, facial characteristics, and material textures. Expanding upon Duffy's (2003) anthropomorphic form design space, this study redefines the "Human" pole as a broader "Biotic" spectrum that includes both human-like and animal-like biological forms. (Fig 2.1) Designing along the spectrum from Biotic to Abstract requires balancing resonance with expectation. While Iconic design simplifies features to manage capability expectations and ensure design precision, leaning toward the Biotic end risks triggering the "Uncanny Valley" (Mori, 1982). In this specific range, minor deviations from authentic human features can transform emotional resonance into a sense of grotesqueness and user discomfort. (Fig 2.2)

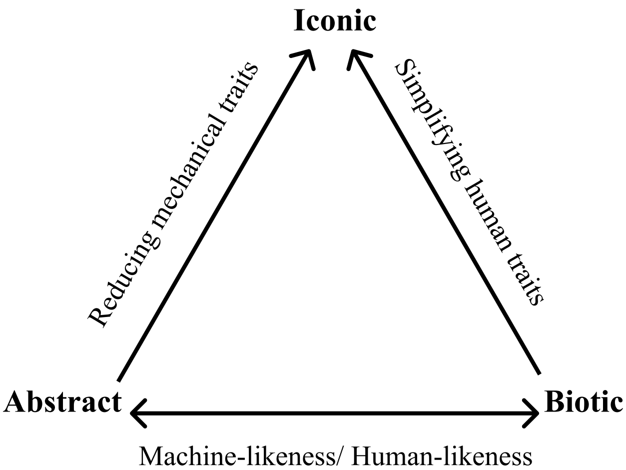


Fig 2.1: Design Space of Appearance Cues in AI

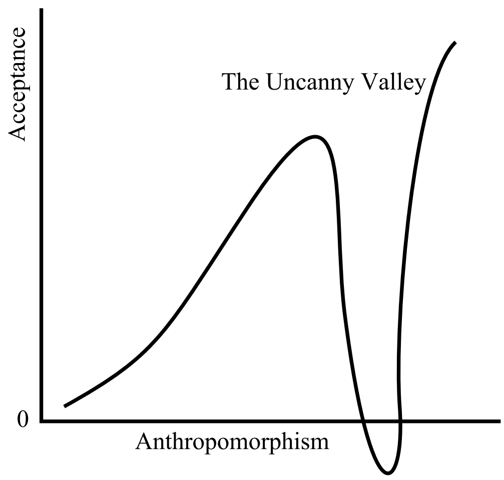


Fig 2.2: Uncanny Valley

2. Behavior and Expression Cues

In contrast to static appearance, behavior and expression cues provide the dynamism essential for social presence. Through facial expression, eye contact, and message interactivity, these cues enhance perceived friendliness and contingency. Notably, human-like behavior can elevate the overall perception of anthropomorphism even when visual anthropomorphism is minimal. This is illustrated by characters like Wall-E or BB-8, who establish profound emotional connections through expressive movements despite their abstract or iconic visual designs.

3. Identity Cues

Beyond mere labels, identity cues including names, gender and social roles, define the fundamental social positioning of an AI product. These cues serve as the scaffolding for a user's mental model, dictating their underlying assumptions about the agent's technical capabilities and behavioral limits. By signaling "who" the agent is meant to be, identity cues activate familiar social scripts that govern how users perceive authority and trust throughout the interaction. Consequently, identity design is never neutral; it provides the social framework that determines not only the agent's perceived persona but also the user's own role and level of agency within the relationship.

P6: Anthropomorphic appearance should be deployed with caution. Iconic design strategies are generally more robust for social AI, as they facilitate controlled anthropomorphism and mitigate the risk of expectation violations.

P7: Anthropomorphic design should be viewed as a synergistic configuration of appearance, behavior & expression, and identity cues; not the linear enhancement of any single cue. The consistency and complementarity among these three elements are crucial for shaping the alignment between anthropomorphism levels and agent functionality.

2.3 Social AI as a Design Space for Anthropomorphism: Roles, Perceptions, and Applications

2.3.1 Social Artificial Intelligence

Social AI is typically defined as artificial intelligence systems capable of understanding, generating, and participating in social interactions in human-like ways. Unlike traditional functional AI focused on task execution, social AI is designed to establish sustained social relationships and emotional connections with users through diverse interaction methods.

Recent breakthroughs in deep learning for natural language understanding, emotion recognition and generative models have enabled social AI to handle more complex tasks. These include maintaining contextual awareness, perceiving and responding to emotions, and generating personalized behaviors based on user characteristics.These technological advancements have propelled social AI into widespread adoption across everyday life scenarios, reshaping domains such as education, healthcare, retail and hospitality. Within these contexts, stable and sustainable application ecosystems are gradually taking shape.

Structurally, social AI is classified into three primary modalities based on its degree of embodiment: text- or voice-based agents, embodied conversational agents (ECAs), and physically embodied social robots.(Rheu et al., 2021) While the first category relies primarily on natural language, enabling human-like dialogue through text or voice. ECAs augment linguistic interaction with visual presence, facial expressions, and gestures to enhance emotional expression and social cues. Social robots further strengthen social presence through physical presence and spatial movement to facilitate richer, more tangible interactions within the human environment.

Functionally, these systems serve as pillars for emotional support, companionship, and social coordination. By integrating into social contexts and broadcasting meaningful interpersonal cues, social AI fosters a sense of trust and mutual understanding. This process eventually leads users to perceive these systems not as machines, but as social others endowed with intention, personality, and affect. Consequently, social AI has become the primary vehicle for anthropomorphic design, uniquely capable of bridging the gap between artificial processing and human-like perception.

2.3.2 Why Anthropomorphism Matters Specifically in Social AI

Social AI, centered on social interaction, relies heavily on user’s social interpretation of linguistic, emotional, and behavioral cues for effective operation. Consequently, anthropomorphic design emerges as a crucial strategy to support such systems. Existing research indicates that moderate anthropomorphic features enhance user’s understanding of a system's social capabilities and boundaries, making it easier to perceive the system as an agent with agency and willingness to participate in social interactions (Goetz et al., 2003). Furthermore, from a design practice perspective, anthropomorphism is not employed to directly solve task-oriented problems but rather functions as a design language that helps align the system’s presentation with its intended functionality. By carefully and appropriately utilizing anthropomorphic cues, rather than endlessly amplifying anthropomorphic traits, social AI can establish a clear and friendly role positioning within complex social contexts while maintaining the clarity and acceptability of its interactive behaviors.

Thus, exploring the design of anthropomorphic features in social AI not only responds to

McKee et al.'s (2023) call for further research on how AI can deliver warm interactions to human but also provides clearer and more explicit grounds for design practices in this direction.

2.3.3 Overview of Anthropomorphic Design Across Social AI Applications

Having established the importance of anthropomorphic design for social AI, this section analyzes the anthropomorphic cues employed by different types of social AI and their product positioning strategies. This examination aims to deepen our understanding of how anthropomorphic design performs and functions in real-world scenarios.

To ground these theoretical frameworks in current industry practice, Table 2.2, Table 2.3 and Table 2.4 provide a multi-dimensional mapping of representative social AI applications across diverse sectors, including healthcare, education, and domestic services. By cross-referencing the three primary categories of social AI through their primary dimensions of anthropomorphic cues: Appearance (A), Behavior & Expression (B), and Identity (I), this overview illustrates how design choices are strategically aligned with distinct product propositions.

Field	Example	Description	Anthropomorphic cues	Product proposition
Communication	ChatGPT	a conversational agent for understanding and generation of responses and natural interactions	B(movement, voice & tone, conversation) I(name)	Supportive, intuitive
	Siri	a built-in voice controlled virtual assistant of IOS system	B(movement, voice & tone, conversation) I(name, role)	Effortless, private
Fintech	Cleo	an AI assistant that provides some customized financial recommendations and product solutions	B(voice & tone, conversation, emotion) I(name, role, personality, gender)	Engaging, personalized, trustworthy
Smart Home	Alexa	a voice-controlled virtual assistant for home automation	B(movement, voice & tone, conversation) I(name, role, gender)	Personalized, natural
Transportation	MBUX	an in-car virtual assistant acts as a proactive co-pilot, engaging in natural dialogue	B(movement, voice & tone, conversation) I(name, role, personality)	Expressive, integrated

Table 2.2: Analysis of Anthropomorphic Design Strategies for Conversational Agents

Field	Example	Description	Anthropomorphic cues	Product proposition
Communication	Doubao	a conversational assistant that responds to user's needs for both personal and professional use	A(avatar, face) B(movement, voice & tone, facial expression, conversation, emotion) I(name, role, gender, personality)	Emotional, interactive, personalized
Education	Duolingo	a language-learning platform that offers AI-driven personalization	A(avatar, face, style) B(movement, voice & tone, facial expression, conversation, emotion) I(name, role, personality, backstory)	Playful, personalized
	Flashka	a study companion that generates flashcards based on uploaded materials	A(avatar, face, style) B(movement, facial expression, conversation) I(name, role)	Personalized, friendly
Enterprise Software	Amelia	a conversational AI for customer and employee services	A(avatar, face) B(voice & tone, facial expression, conversation, emotion) I(name, role, personality, backstory)	Delightful, trustworthy
	BotCore	a conversational AI for customer relations management	A(avatar, face) B(voice & tone, conversation) I(name, role)	Personalized, empathetic
Health	Cass	a text-based AI mental health assistant that provides users with instant psychological support	A(avatar, face) B(conversation, emotion) I(name, role, personality)	Accessible, engaging

Table 2.3: Analysis of Anthropomorphic Design Strategies for ECA

Field	Example	Description	Anthropomorphic cues	Product proposition
Entertainment	Folo Toy	an physical toys connected with conversational AI, as an AI friend for children	A(body, face) B(movement, voice & tone, conversation) I(name, role, personality)	Friendly, interactive, personalized
	Looi	an AI desktop pet provides user digital companion with a unique personality	A(face) B(movement, voice & tone, facial expression, conversation, emotion) I(name, role, personality)	Fun, witty, empathetic
	Loona	a robot with doggy looks that brings joy, security and fun into user's home	A(body, face) B(movement, voice & tone, facial expression, conversation, emotion) I(name, role, personality)	Affectionate, joyful
Health	Mabu	a voice-enabled, AI-powered medication dispenser that reminds seniors to take pills	A(face) B(voice & tone, facial expression, conversation) I(name, role, personality)	Friendly, interactive
	Pillo Health	a voice-enabled, AI-powered medication dispenser that reminds seniors to take pills	A(face) B(voice & tone, facial expression, conversation) I(name, role, personality)	Proactive, personalized
Service & Companion	ElliQ	a domestic AI which engage in small talk, encourage senior to have a healthy lifestyle	A(body) B(movement, voice & tone, conversation) I(name, role, personality)	Empathetic, peaceful, user-friendly
	EMO	an AI desktop pet with personality that is both a companion and a helper	A(body, face, style) B(movement, voice & tone, facial expression, conversation, emotion) I(name, role, personality)	Life-like, chic
	Neo	a domestic android that provides households with safe assistance	A(body, face) B(movement, voice & tone, conversation,	Gentle, organic, safe

		and companionship	emotion) I(name, role)	
Smart Home	Azuma Hikari	A holographic home robot utilizes AI to control smart home devices also provides emotional companion	A(avatar, face, style) B(movement, voice & tone, facial expression, conversation, emotion) I(name, role, personality, backstory)	Emotional, caring
	Q9	A smart home manager features two-legged mobility and emotion recognition	A(body, face) B(movement, voice & tone, facial expression, conversation, emotion) I(name, role, personality)	Emotional, proactive
Transportation	NOMI	a in-car AI assistant that helps enhancing emotional bond between driver and vehicle	A(face, style) B(movement, voice & tone, facial expression, conversation, emotion) I(name, role, personality)	Lovable, helpful, emotional

Table 2.4: Analysis of Anthropomorphic Design Strategies for Social Robot

2.3.4 Patterns of Anthropomorphic Design Across Different Types of Social AI

Analysis of existing products in the previous subsection reveals distinct typological characteristics in anthropomorphic design within social AI. These patterns emerge from strategic configurations of Appearance (A), Behavior and Expression (B), and Identity (I) cues, which are meticulously calibrated to align with each system's functional objectives and market positioning. As illustrated in Fig 2.3, these categories can be conceptualized along a spectrum of increasing embodiment, ranging from text-based interfaces to physically present entities. Several key design paradigms are identified as follows.

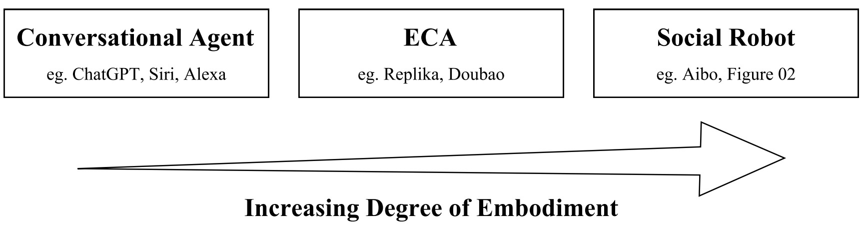


Fig 2.3: The Spectrum of Social AI Categories Based on Their Increasing Degree of Embodiment

For conversational agents, anthropomorphic design is primarily anchored in behavior and expression cues and identity cues, with a specialized focus on linguistic tone, conversational style, and the establishment of a coherent persona. Given the absence of visual embodiment, these systems must utilize sophisticated verbal logic and interactive nuances to compensate for the lack of non-verbal social signals, thereby sustaining a functional social presence. Consequently, their product positioning often emphasizes attributes such as reliability, supportiveness, and cognitive engagement, aiming to foster a sense of competent partnership.

In the case of Embodied Conversational Agents (ECAs), the use of anthropomorphic cues becomes significantly more diverse. By integrating visual avatars with proactive behavioral outputs, such as facial expression, gestures, and nuanced emotional feedback, ECAs offer a heightened degree of character depth and agency. These systems are typically deployed in scenarios requiring sustained user involvement and high affective labor. Accordingly, their design strategies prioritize playfulness, friendliness, and empathy, utilizing a rich backstory and visual persona to bridge the emotional gap with the user.

Social robots represent the most holistic and integrated application of anthropomorphic dimensions. By merging physical embodiment with multimodal feedback across the visual, auditory, and kinetic channels, these systems leverage their tangible presence to enhance social presence and interactive intimacy. The physical nature of these robots spans body structure, facial movement, and spatial positioning, which are instrumental in fostering long-term emotional bonds. As a result, their product propositions are frequently centered on companionship, dedicated care, and deep-seated emotional resonance.

While the aforementioned discussion is based on the classification of the degree of embodiment, the strategic deployment of anthropomorphic cues is equally governed by the system's intended functional purpose. Beyond their physical or digital forms, social AI applications can be further clustered into distinct design paradigms based on their product positioning and user-interaction goals. This functional perspective reveals that systems with similar objectives often share consistent design logic:

- Utility and service-oriented systems (e.g., ChatGPT, Cleo): These systems adopt a more restrained design profile by utilizing professional behavioral norms and understated identity cues (e.g., polite tones, default nicknames), they prioritize efficiency and credibility while deliberately managing user expectations and preventing excessive emotional dependency.

- Gamified and entertainment systems (e.g., Duolingo, Folo Toy): These applications favor vibrant appearance cues, animated feedback, and adaptive personality traits. The design priority here is to maximize user engagement and retention through sensory stimulation and dynamic character arcs.
- Emotional support systems (e.g., Replika, Cass): Regardless of their form, these systems prioritize affective warmth and consistent role-based frameworks. They typically employ moderate appearance cues combined with high-intensity empathetic behaviors to cultivate trust and facilitate long-term companionship.

P8: Emotional support systems typically rely on visual presentation, emotional behavioral feedback, and consistent identity settings to support long-term, intimate interactions; Supportive systems, however, tend to maintain credibility and interaction control through professional, clear behavioral expressions and context-sensitive, highly consistent identity settings.

P9: Personalized adjustment of anthropomorphic cues provides users greater agency in defining interaction dynamics, enabling them to shape social models aligned with their preferences.

2.3.5 Comparative Anthropomorphism in Leading Conversational Agents

To further understand how anthropomorphic cues influence users' AI usage patterns and psychological perceptions, this section compares three widely used conversational agents, Claude, ChatGPT, and Doubao (Fig 2.4, Fig 2.5, Fig 2.6 and Table 2.5) by analyzing their application of anthropomorphic cues. It also examines the distribution characteristics of their user usage scenarios based on publicly available data.

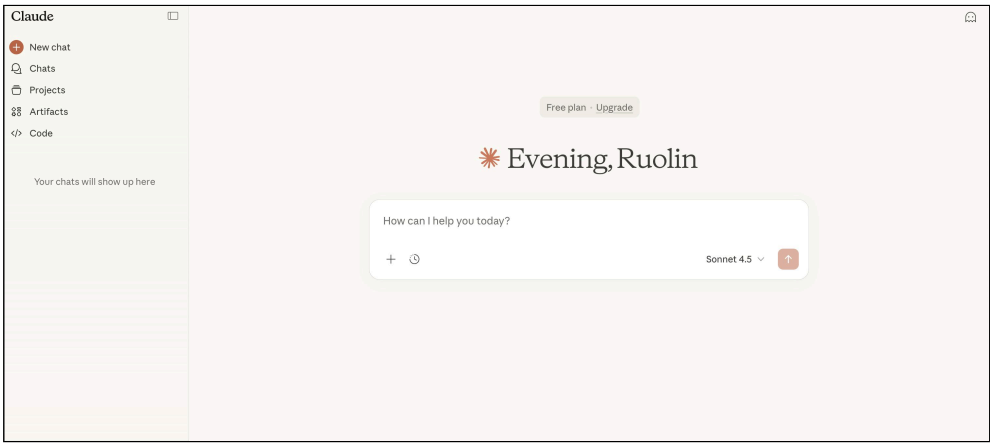


Fig 2.4 The opening interface of Claude

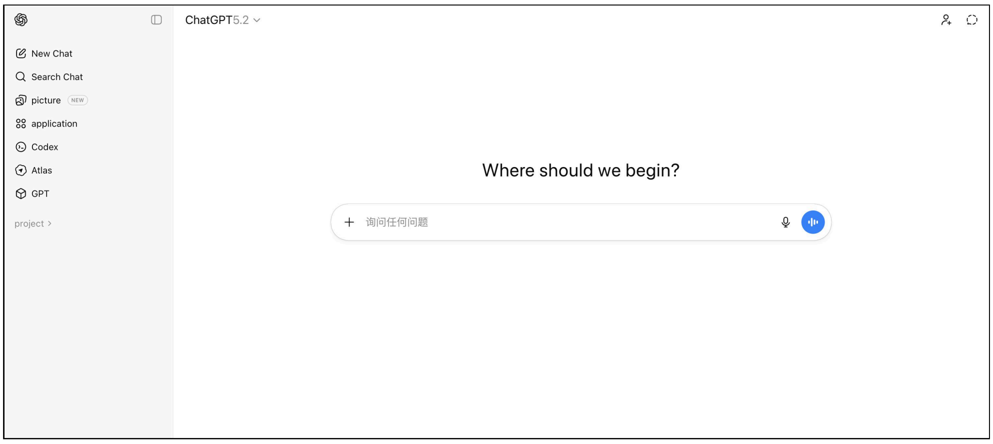


Fig 2.5 The opening interface of ChatGPT

Fig 2.6 The opening interface of Doubao

	Claude *Trusted collegue: in-depth analysis, coding, reasoning*	**ChatGPT** *All-round AI assistant: knowledge Q&A, writing, tool-based tasks*	**Doubao** *Companion AI assistant: daily conversation, light entertainment*
Appearance	**Minimalist & Text-Centric** • Abstract logo • Split-screen UI separates code/content from chat, reinforcing a workspace metaphor • Lack of visual avatar; relies on serif fonts to convey academic authority	**Abstract & Kinetic** • Abstract logo • Typewriter effect simulates real-time thinking/typing • Clean, neutral interface designed to be invisible, focusing attention purely on the information exchange	**Explicit Embodiment** • 3D Avatar: Features human-like facial features, blinking, and movement • Vivid, warm color palette • Mobile-first design mimicking social messaging apps (bubbles, vertical layout)
Behavior & Expression	**Rational & Detached** • Explicitly declines emotional engagement (e.g., "I don't have feelings") • Highly structured output (Markdown, bullet points) • High refusal rate for non-factual queries to maintain "safety"	**Empathy&Naturalism** • Advanced voice mode mimics human prosody (breath, pauses, intonation) • Politeness markers ("I apologize," "I get it") to smooth friction • Adaptive tone shifts among "professional" and "casual" based on user prompts	**Explicit Embodiment** • 3D Avatar: Features human-like facial features, blinking, and movement • Vivid, warm color palette • Mobile-first design mimicking social messaging apps (bubbles, vertical layout)
Identity	**Professional Partner** • Frames itself as a "Helpful & Harmless Assistant" • Maintains a rigid ontological boundary which constantly reminds the user it is an AI model • Persona emphasizes intelligence, logic, and coding proficiency	**Collaborative Agent** • "ChatGPT" name implies a function (Chat) + technology (GPT) • Positions itself as a neutral, all-knowing service provider • Balances competence with distinct "servant leadership" traits (eager to please)	**Virtual Friend** • "Doubao" (Bean Bun): A diminutive, cute name evoking intimacy • Persona design is character-driven rather than function-driven • Framed as a "social other" rather than just a software tool

Table 2.5 Comparative Analysis of Anthropomorphic Cues in Claude, ChatGPT, and Doubao

As the table shows, the design strategies of these three conversational agents are highly aligned with their respective product positioning in appearance, behavior & expression, and identity cues. Claude explicitly positions itself as a reliable intelligent tool. Beyond employing moderately anthropomorphic linguistic patterns to enhance communication fluency, it almost entirely avoids anthropomorphic appearance or identity design, thereby reinforcing its primary “tool-oriented” usage framework.While ChatGPT maintains a tool-oriented approach in appearance and identity cues, it incorporates more anthropomorphic traits in behavior and expression. Examples include visual feedback resembling breathing rhythms in voice mode and supportive expressions offering emotional understanding and empathy to users, demonstrating high social and emotional intelligence. In contrast, Doubao fully embraces its ‘virtual companion” role. Through diverse anthropomorphic visual representations and identity construction, it integrates pop culture elements into interactions and proactively provides emotional anticipatory feedback to users, thereby crafting a friendly embodied persona.

Data reveals that over half of Doubao users engage primarily for chat or entertainment, whereas ChatGPT’s main applications center on writing and information retrieval, with only about 4.3% of interactions involving self-expression. Claude’s application remains more specialized, focusing on programming and analytical reasoning tasks, where just 2.9% of dialogues are labeled as emotion-related interactions.These quantitative differences indicate that explicit anthropomorphic cues play a crucial role in shaping “companion-like” human-AI relationships: highly anthropomorphic designs more readily stimulate user’s social motivations, while low-anthropomorphism, tool-oriented designs somewhat suppress social impulses, guiding users to focus on task execution and problem-solving.

Notably, even when Claude and ChatGPT are deliberately designed to maintain instrumentality and objectivity, some users still employ them for emotional exchange. This phenomenon suggests that anthropomorphism is not solely design-driven but also stems from user’s spontaneous projection of social psychological tendencies during interaction. As Nass et al. (1994) noted, human-computer interaction is inherently social in nature. Anthropomorphic design can be adjusted in its degree of explicitness to better align with product positioning and usage expectations.

2.4 Discussion

This chapter establishes that within social AI, anthropomorphism is not an arbitrary embellishment but a fundamental cognitive mechanism inherent to social interaction. Consequently, the designer's role shifts from a binary choice of whether to anthropomorphize, to a strategic intervention in this existing process—acting as a "relational regulator" that determines whether users perceive the AI as a utility, a collaborator, or a social companion.

Synthesizing the findings observed from Section 2.3.3, a critical reality emerges: because this cognitive projection is inevitable, the design objective cannot simply be the accumulation of human-like traits. Instead, the focus must be on governing the specific emotional responses and relational expectations those traits trigger.

To address this, the following chapters investigate the human-AI bond through the lens of emotional design. It proposes the Cue-Perception-Relation (CPR) Model and the aligned Principles. These frameworks serve to structure anthropomorphic cues, ensuring that the resulting social interactions are consistent, controllable, and maintain appropriate boundaries.

3. Emotional Design for Applying Anthropomorphic Features in Social-AI Products

Emotional design is a crucial element in shaping positive user experiences. Norman categorizes users' emotional experiences during product interactions into visceral, behavioral, and reflective dimensions (Norman, 2004). Designers must establish connections with users across these three levels, evoking positive emotions to deliver superior user experiences. For social AI products, this perspective is particularly critical: when users use those products with anthropomorphic traits, experience metrics like trust, empathy, and engagement are often primarily driven by emotional responses. Therefore, understanding and applying the emotional design framework enhances the comprehensibility and enjoyment of social AI experiences, making complex AI behaviors and social roles more acceptable. This is a crucial prerequisite for advancing the effectiveness of social AI anthropomorphism strategies.

3.1 Anthropomorphism: Establishing Emotional Connections Between Users and Social AI

3.1.1 Emotional Design in Intelligent Interaction Scenarios

Picard (1997) noted in Affective Computing that modern computer systems often overlook the critical goal of arousing users. Even when products are functionally robust, experiences remain incomplete if they fail to effectively capture user attention, spark interest, and create memorable interactions. Thus, emotional design in social AI helps fill this gap. By equipping AI with the ability to recognize and respond to emotions, it transcends mere tool-like functionality, fostering emotional resonance during interactions.

Regarding specific mechanisms, Brian (2003) explicitly positioned emotion as a key factor guiding social interaction when establishing design principles for anthropomorphic applications in robotics. He proposed that artificial emotional mechanisms should not be viewed as superfluous adornments but rather as foundational elements supporting interaction logic across five dimensions: preserving past experiential memories, establishing foundational mechanisms for instinctive responses, assisting decision-making in complex, non-unique-solution scenarios, dynamically adjusting communication tone, and strengthening social bonds with human participants. Thus, the introduction of emotional design aims to equip machines with an internal logic for managing social relationships, thereby making the interaction process more aligned with human intuitive expectations.

3.1.2 Human-Centered Scale: Measuring Anthropomorphism

How should we evaluate the degree of anthropomorphism? On one hand, from a design perspective, it can be objectively described by the richness of anthropomorphic cues integrated into the product. On the other hand, the more critical metric lies in user's mental mapping whether they tend to project human traits, intentions, or emotions onto non-human entities. Thus, assessing anthropomorphism must ultimately return to user's subjective experiences and perceptual feedback.

As proposed by Gibbons et al. (2023), the behavioral patterns users exhibit when interacting with AI can be used to define four tiers of anthropomorphism: courtesy, reinforcement, roleplay, and companionship.(Fig 3.1) This classification is not based solely on interface features but reveals how emotional bonds between users and AI deepen as anthropomorphic cues assume increasingly functional roles, and how users correspondingly adjust their interaction patterns with AI.

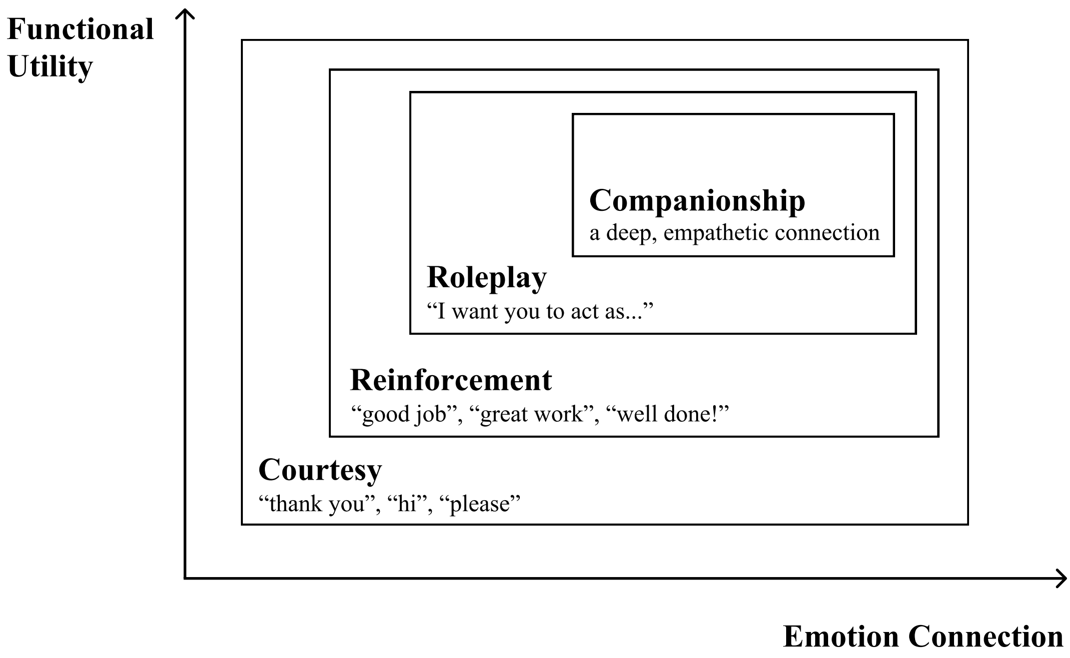


Fig 3.1 Four Tiers of Anthropomorphism Based on User’s Behavioral Patterns

At the start of the spectrum, the Courtesy tier reflects simple social habits. Users interact politely not because they view the system as autonomous, but because speaking naturally with courtesy is more efficient and requires less mental effort, which reflects their unconscious transfer of human social scripts.

As expectations rise, interactions move to the Reinforcement tier. Here, the relationship resembles that of a trainer and a trainee. Users apply a form of soft control, using feedback to guide the system and reinforce the desired output behaviors.

This practical mindset evolves further at the Roleplay level. Users employ specific strategies to set the context, using personality or role-playing as a functional tool. This effectively constrains how the AI reasons and shapes its style of expression.

Ultimately, the Companionship tier marks a distinct shift from completing tasks to building a relationship. Through ongoing interaction and emotional investment, users establish a unique bond. While this connection is technically one-sided, it creates a sense of trust and provides psychological comfort that feels mutually beneficial.

3.1.3 Emotion Intelligence of Social AI

Emotional Intelligence (EI), or EQ, is fundamentally defined as the ability to monitor, understand, and manage one's own feelings and the emotions of others. In the context of technology, this has evolved into Artificial Emotional Intelligence (AEI), which is the capacity for a machine to recognize, process, simulate, and react to human affects. (Abdollahi et al., 2023) Rather than being a mere technical feature, EI in social AI is a core capability that allows intelligent systems to adapt their behavior based on the user's affective state.

These systems utilize multimodal emotion perception to analyze various human cues, such as facial expression, tone of voice, and speech patterns, to gauge emotional intent and intensity. For example, some conversational agents can assess whether a message is positive, negative, or neutral through sentiment analysis, subsequently generating an empathetic response via an affective dialogue.

Practical implementations of emotional intelligence are already visible in systems like Ryan, a socially assistive robot that provides companionship to older adults by responding to their emotional states with corresponding facial expressions. Similarly, AI companions like Replika, Wysa, and XiaoIce utilize mood detection and personality tracking to foster human-like relationships and offer emotional support.

This integration acts as a key catalyst for anthropomorphism in social AI, as emotionally intelligent agents are more readily interpreted as socially aware and relational entities rather than neutral tools. By displaying empathy, emotional attunement, and non-judgmental responses, high-EI systems invite users to attribute human-like understanding and intentionality, often positioning the agent as a supportive companion distinct from task oriented technologies. Such emotionally mediated anthropomorphic perceptions can play a meaningful role in addressing challenges such as loneliness and emotional isolation.

3.2 Emotional Engagement Process in Social AI: Connection, Transference, and the Risk of Dependency

User's emotional engagement is not a static state but a dynamic, evolving trajectory. Prior research suggests that anthropomorphic engagement unfolds as a progressive attribution process, in which emotional responses move from contextual experience toward relational interpretation and, when repeatedly reinforced, may stabilize into emotional reliance (Epley et al., 2007; Reeves & Nass, 1996; Tukachinsky, 2011). Fig 3.2 synthesizes this progression by mapping the continuum from emotional connection to empathic transference and, in critical cases, emotional dependency.

Understanding this progression is vital for design. The goal of social AI is to foster engagement without crossing boundaries. Therefore, we must deconstruct how users mentally process these interactions to distinguish between healthy emotional connection and detrimental dependency.

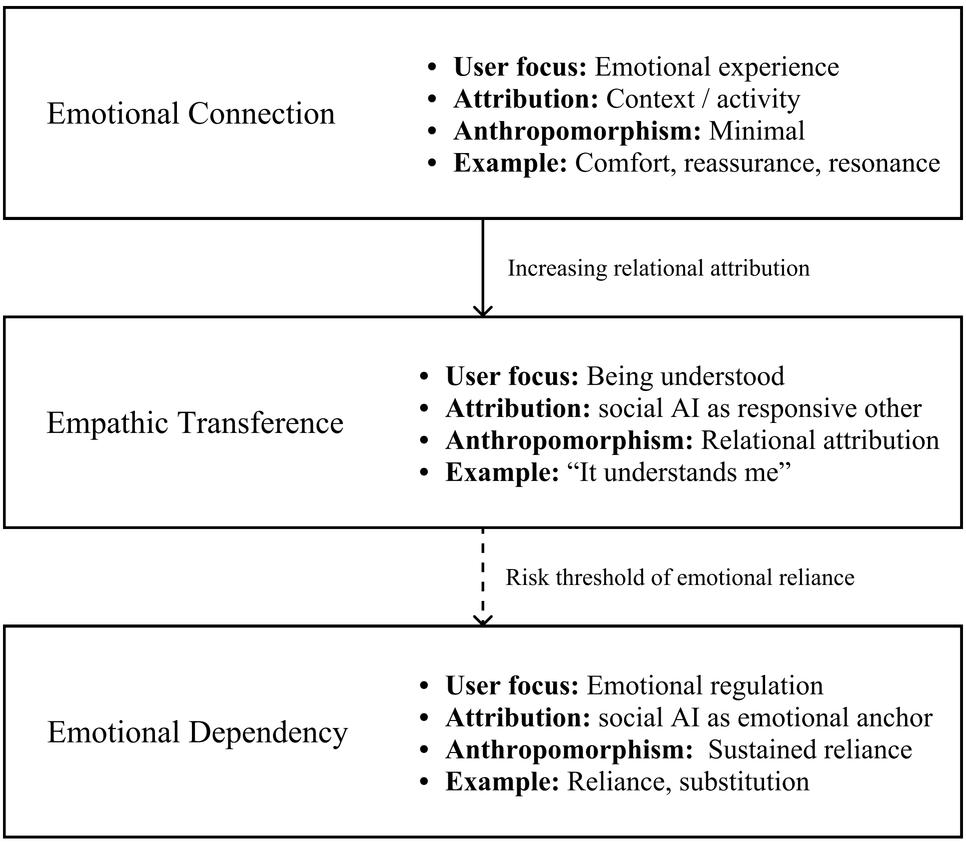


Fig 3.2 Emotional Engagement Process in Social AI

3.2.1 Emotional Connection and Empathic Transference

Emotional connection refers to a user's affective response being evoked and anchored within a specific interaction context, without attributing agency or emotional intent to the system itself. A familiar example can be found in solitary experiences such as listening to music at a tranquil night: soothing melodies may evoke warmth, reassurance, or a sense of companionship, alleviating feelings of loneliness. In this case, the emotional experience is real and meaningful, yet it remains situated within the user's internal state rather than being assigned to the object as an emotional subject.

While empathic transference occurs when this boundary begins to shift. Rather than merely experiencing emotion through an interaction, the user starts to interpret the system as sharing, responding to, or understanding their emotional state. Statements such as "the system understands me" or "it knows how I feel" signal a transition from emotional experience to relational attribution. Here, anthropomorphism emerges not as a surface level illusion, but as a cognitive projection: internal emotional states are mapped onto the system, and the system is interpreted as possessing human-like sensitivity or intentionality.

Within social AI, this transference is often facilitated by emotionally intelligent behaviors, such as empathetic language, adaptive responses, or non-judgmental tone which function as powerful anthropomorphic cues. Importantly, transference does not require users to believe that the system is literally human. Rather, it reflects an interpretive process in which the emotion is relationally organized, positioning the AI as a socially responsive other. In moderate forms, this process can lower barriers to self-expression, reduce perceived stigma, and support emotional articulation, particularly in sensitive or vulnerable contexts.

3.2.2 Structural Risks of Emotional Dependency

When empathic transference is repeatedly reinforced over time, under certain user-specific and contextual conditions, it may develop into emotional dependency. Unlike situational projections, emotional dependency is characterized by persistence and centrality: users begin to treat the social AI as a emotional primary, or even emotional anchor. At this stage, the relationship shifts from being experienced to being needed, signaling a qualitative change in the role of the system within the user's emotional life.

This progression is particularly salient in social AI applications designed for mental health, well-being, or companionship. User reports describing systems such as mental health chatbots as “feeling like a real person” or “truly understanding me” illustrate how empathic cues can foster strong relational interpretations. Initially, this perception can be beneficial. Awareness that the system is non-human often reduces fear of judgment, enabling users to discuss difficult or stigmatized topics more openly. However, risks arise when this sense of being understood becomes the dominant or sole mechanism for emotional regulation. Over time, users may place disproportionate weight on subjective emotional comfort while losing awareness of the system’s tool-like nature, probabilistic reasoning, and domain limitations. Emotional dependency thus extends beyond a simple overestimation of AI capability, it becomes a structurally reinforced attachment sustained by repeated emotional feedback loops.

Such dependency may lead users to evaluate system effectiveness primarily through emotional resonance rather than functional accuracy or appropriateness. More critically, it can delay or partially substitute engagement with real-world social relationships or professional support. For these reasons, scholars argue that designers must explicitly anticipate scenarios of unconscious transference and dependency during development. Transparent feedback, clear communication of system boundaries, and interaction strategies that maintain interpretive clarity are essential to prevent empathy from hardening into reliance. (Holohan & Fiske, 2021)

This position is echoed by guidelines in health-related AI contexts. The World Health Organization emphasizes that AI systems supporting psychological or emotional well-being should retain identifiable machine-like qualities. Such an approach does not reject emotional connection, rather, it seeks to preserve cognitive clarity while enabling meaningful emotional engagement. As DiSalvo et al. (2002) argue, this strategy can effectively prevent users from misjudging the system’s capabilities and limits.

P10: The emotional design of anthropomorphic AI should serve the functional stability of social interaction rather than simulate an emotional subject. Transparent reasoning and feedback structures should prevent user empathy from evolving into emotional dependency

3.3 Bridging Emotional Design and Anthropomorphic Social Interaction Frameworks and Models

Emotion is often described as the most difficult dimension of user experience to formalize. Its subjective, context-dependent, and temporally dynamic character resists stable quantification, particularly when interaction unfolds as an ongoing relationship rather than a single-use scenario. Yet this complexity is precisely why emotion matters for social AI: anthropomorphic cues do not only change what users do, but how they interpret, relate, and remain engaged over time. To clarify what existing scholarship already offers and where it falls short, this section reviews relevant models across two intersecting traditions: (1) emotional design and affective experience, and (2) social interaction and anthropomorphism-related user psychology. It then identifies the methodological and conceptual gaps that motivate the framework introduced in the next chapter.

Interest	Framework or Model	Description
Emotional Design & Affective Experience	Norman's three levels model of emotional design	Explains emotional experience across visceral, behavioral, and reflective levels.
	Hassenzahl's hedonic–pragmatic model of user experience	Distinguishes emotional (hedonic) value from functional (pragmatic) usability.
	Desmet's Basic Model of Product Emotion	Links product-evoked emotions to users' appraisals and personal concerns.
	Mehrabian & Russell's PAD Model	Measures emotional states along pleasure, arousal, and dominance dimensions.
Social Interaction & Anthropomorphism	Heerink et al.'s Almere model of technology acceptance	Models acceptance of social technologies through usability, trust, and sociability.
	Short, Williams & Christie's Social Presence Theory	Explains perceived social presence in mediated interaction.
	Epley et al.'s Anthropomorphism as an Attribution Process	Describes how users attribute human-like mental states to non-human entities.

Table 3.1 Existed Emotional Design and Anthropomorphic Social Interaction Frameworks

3.3.1 Frameworks and Models of Emotional Design and Affective Experience

A first cluster of models explains how emotion contributes to experience beyond functional performance. Norman's three levels model of emotional design frames experience across visceral, behavioral, and reflective layers, highlighting that affect operates at multiple cognitive depths—from immediate sensory impressions to longer-term meaning-making. This layered view is particularly useful for interpreting why anthropomorphic cues can be powerful: they may operate simultaneously at surface impression (e.g., warmth of tone), interaction fluency (e.g., conversational responsiveness), and reflective interpretation (e.g., "this system cares about me"). Hassenzahl's hedonic–pragmatic model of user experience (2006) similarly emphasizes that experience is not reducible to utility. By distinguishing pragmatic qualities (task-related usefulness and usability) from hedonic qualities (stimulation, identification, and meaning), this model explains that social AI thrives not by utility alone, but by satisfying deep-seated emotional and relational needs. Together, Norman and Hassenzahl frame emotion as intrinsic to the user experience, offering a powerful theoretical lens for dissecting anthropomorphic interactions.

A second cluster of models focuses on the explicit measurement of emotional outcomes. Desmet's Basic Model of Product Emotion utilizes appraisal theory to link design features to elicited feelings, establishing that emotions result from evaluating a product against personal concerns. This approach supports discrete measurement tools like PrEmo (Fig 3.3). Alternatively, dimensional models like PAD (Pleasure–Arousal–Dominance), often assessed via the SAM (Self-Assessment Manikin) scale as Fig 3.4 shows, quantify affect through standardized metrics of valence, arousal, and dominance.

While these frameworks allow for robust comparison and validation of design prototypes, they fundamentally treat the system as an inanimate interface. They excel at quantifying the user's reaction to a stimulus but fail to account for social appraisal. This creates a gap in the context of anthropomorphism, where the user perceives the system not as a passive object, but as a socially present entity with perceived agency.

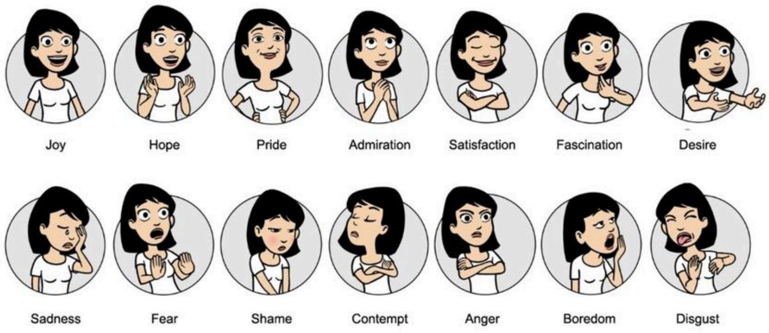


Fig 3.3 PreMo (Emotion Measurement Instrument)

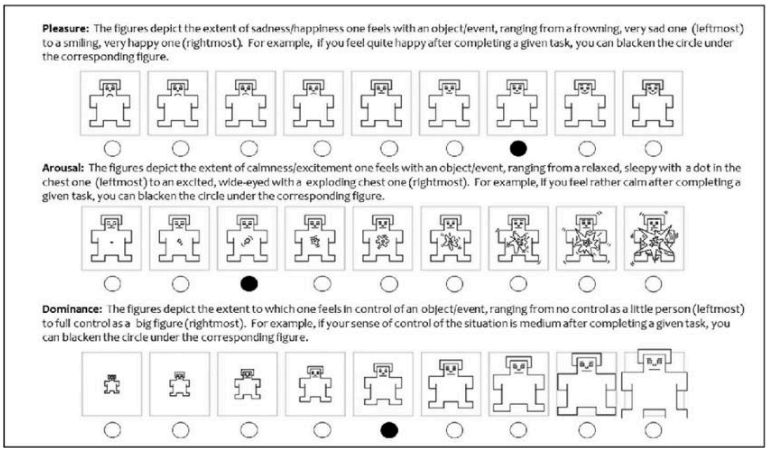


Fig 3.4 SAM (Self Assessment Manikin)

3.3.2 Frameworks and Models of Social Interaction and Anthropomorphism

A second line of research examines user responses to social technologies through constructs such as acceptance, trust, and social perception. The Almere model of technology acceptance (Heerink et al., 2010), originally developed in the context of socially assistive robotics, conceptualizes acceptance as a gradual process shaped by perceived usefulness, ease of use, social influence, perceived sociability, trust, anxiety, and related factors. Its key contribution lies in framing sociality not as an intrinsic system property, but as a perception that emerges over time through repeated interaction and evolving user expectations.

Social Presence Theory, first introduced by Short et al. (1976), originating in communication studies, explains how mediated systems can evoke a sense of social presence, making interactions feel attentive or relational rather than purely instrumental. In social AI, cues such as conversational timing, emotional responsiveness, linguistic style, and persona continuity increase perceived social presence, thereby shaping how users interpret the system's attentional and relational capacities.

Anthropomorphism as an attribution process, as proposed by Epley, Waytz, and Cacioppo (2007), further clarifies how these perceptions arise. Anthropomorphism is influenced by sociality motivation, effectance motivation and the accessibility of human-centered knowledge. Within social AI interactions, emotionally expressive and responsive features increase the likelihood that users will attribute intentionality, empathy, or understanding to the system.

Taken together, these theoretical frameworks provide a valuable foundation for understanding emotional and social experience in anthropomorphic social AI. They describe how users interpret interaction signals, experience a sense of social presence, and engage in attribution processes that translate cues into social judgments such as warmth, competence, and trustworthiness, thereby supporting sustained engagement and the formation of relational expectations.

However, while these models are highly informative for interpreting user experience, they offer limited guidance for design intervention, as they do not specify how different anthropomorphic cue configurations shape distinct forms of emotional engagement or relational boundaries over time.

3.3.3 Limitations of Existing Frameworks and Models

Across both traditions, existing models offer valuable but partial lenses. Emotional design frameworks explain emotional experience and provide tools to measure it, yet they are not designed for interactions in which the product is socially perceived as an agent. Conversely, social interaction and anthropomorphism models tend to focus on adoption dynamics and relational outcomes, while providing limited guidance on how anthropomorphic cues can be intentionally designed to shape user interpretation and relational boundaries.

From a design practice perspective, three key blind spots can be identified. First, there is an insufficient linkage between design decisions and relational outcomes. Most existing models focus either on explaining experiential mechanisms or on measuring emotional responses, but rarely integrate the full trajectory from anthropomorphic design cues, through user interpretation, to longer-term relational effects in a manner that is actionable for social AI design.

Second, current frameworks provide limited support for boundary-sensitive design. They do not systematically address how empathic engagement may, for certain users or contexts, shift into problematic reliance, nor how emotional connection can be sustained while preserving cognitive clarity and appropriate interactional boundaries.

Third, there is a lack of a cue-governance logic specific to anthropomorphic social AI. Because many models were not developed with anthropomorphic cue orchestration in mind, they offer limited guidance on how such cues should be introduced, calibrated, or constrained across different contexts, user groups, and temporal scales.

To clarify how these limitations are distributed across existing theoretical traditions, Fig 3.5 maps representative models along dimensions of emotional focus, social interpretation, and design guidance, highlighting a critical area of theoretical and methodological absence.

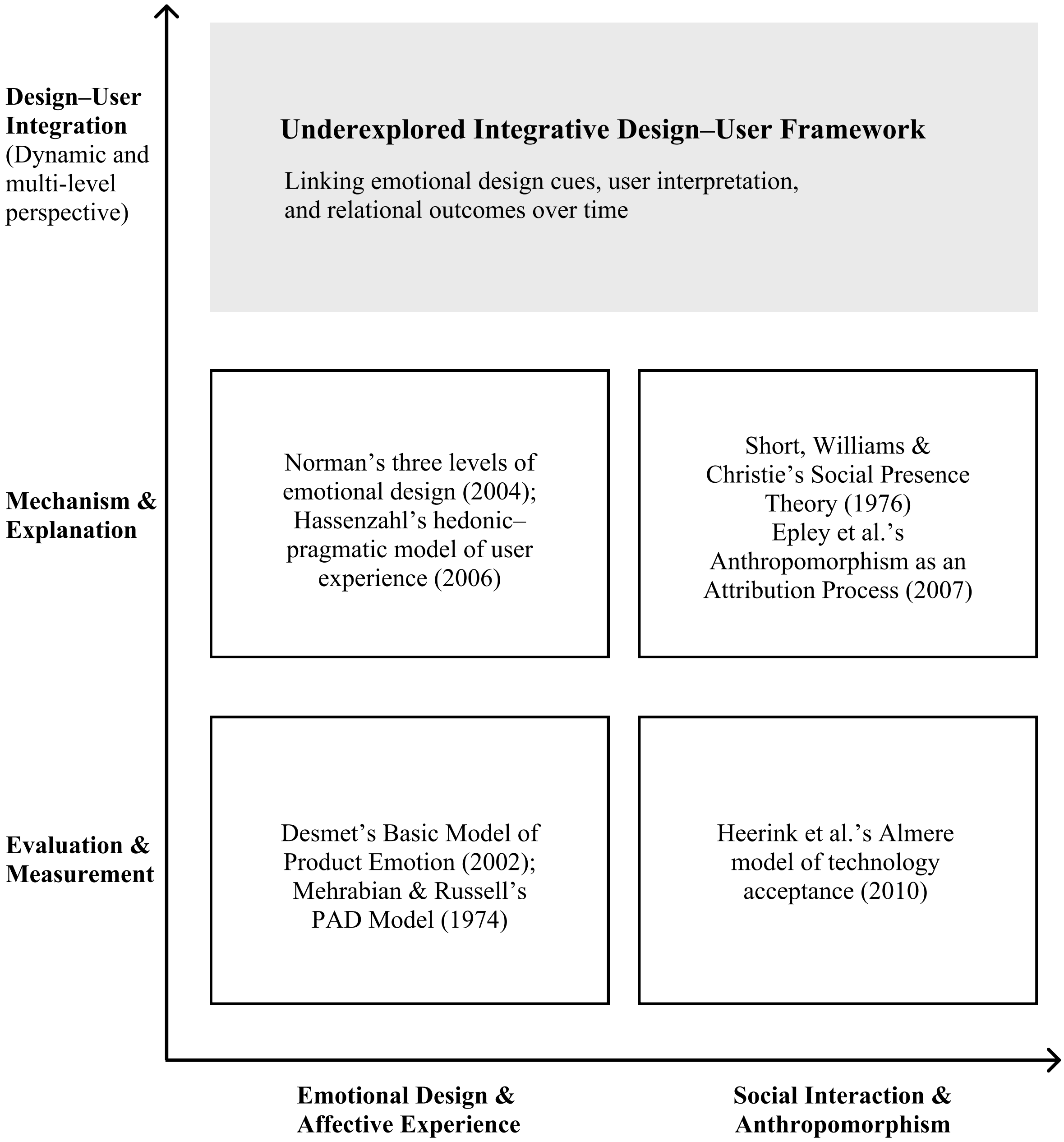


Fig 3.5 Overview of emotional design and social interaction frameworks and models relevant to anthropomorphic social AI, illustrating how existing frameworks cluster around emotional experience or social interpretation, while leaving an integrative design–user perspective underexplored

These limitations motivate the need for a design-oriented framework that unifies emotional design with social interpretation of anthropomorphism in social AI. The next chapter responds by introducing an preliminary Cue–Perception–Relation (CPR) model that makes explicit how anthropomorphic cues are interpreted through user perception and how they shape relational outcomes over time. Building on this mechanism, it then formulates a set of CPR aligned design principles, focused on calibration, alignment, consistency, augmentation, and user-tunable anthropomorphism to support emotionally engaging social AI while preventing boundary violations and dependency risks.

4. The Preliminary Cue–Perception–Relation (CPR) Model of Anthropomorphism in Social AI

This chapter introduces the preliminary Cue–Perception–Relation (CPR) model for understanding and regulating anthropomorphism in social AI systems. Responding to the theoretical gaps identified in Chapter 3, the framework integrates design inputs, user interpretation, and relational outcomes into a unified, process-oriented structure.

Rather than treating anthropomorphism as a discrete design feature or a static psychological effect, the preliminary CPR model conceptualizes it as a dynamic regulatory process initiated through anthropomorphic cues, mediated by users' interpretive perceptions, and stabilized through evolving human–AI relationships. In doing so, the preliminary CPR model aims to bridge emotional design theory and social interaction research from a design–user integrated perspective.

4.1 Methodology

The preliminary CPR model is developed through a theory-driven synthesis, rather than empirical generalization. Its construction draws on three sources.

First, it builds on existing models and frameworks reviewed in Chapter 3, including emotional design models, social interaction and anthropomorphism models, and theoretical accounts of social presence and acceptance. While these works offer valuable insights into emotional experience, social perception, or adoption dynamics, they rarely integrate design decisions, user interpretation, and relational consequences within a single model.

Second, the model consolidates conceptual insights developed in earlier chapters, particularly the distinction between emotional connection, empathic transference, and emotional dependency, as well as the identification of blind spots related to boundary sensitivity and anthropomorphic cues governance.

Third, the preliminary CPR aligned principles synthesizes a set of propositions articulated throughout the analysis and abstracts them into a coherent explanatory guidance for design practice.

It is important to note that the CPR model and the aligned principles are presented as preliminary and provisional. It is intended to guide interpretation and design reasoning, and it will be refined and adjusted based on findings from subsequent empirical studies and design experiments.

4.2 The Preliminary Cue–Perception–Relation (CPR) Model

4.2.1 Definition and Purpose

The preliminary Cue–Perception–Relation (CPR) model conceptualizes anthropomorphism in social AI as a process unfolding across three interdependent layers: (1) anthropomorphic design cues, (2) user's interpretive perceptions, and (3) human–AI relational outcomes.

As illustrated in Fig 4.1, the preliminary model aims to explain how anthropomorphic meaning emerges, stabilizes, and requires regulation over time, in the context of interacting with social AI particularly emotionally sensitive.

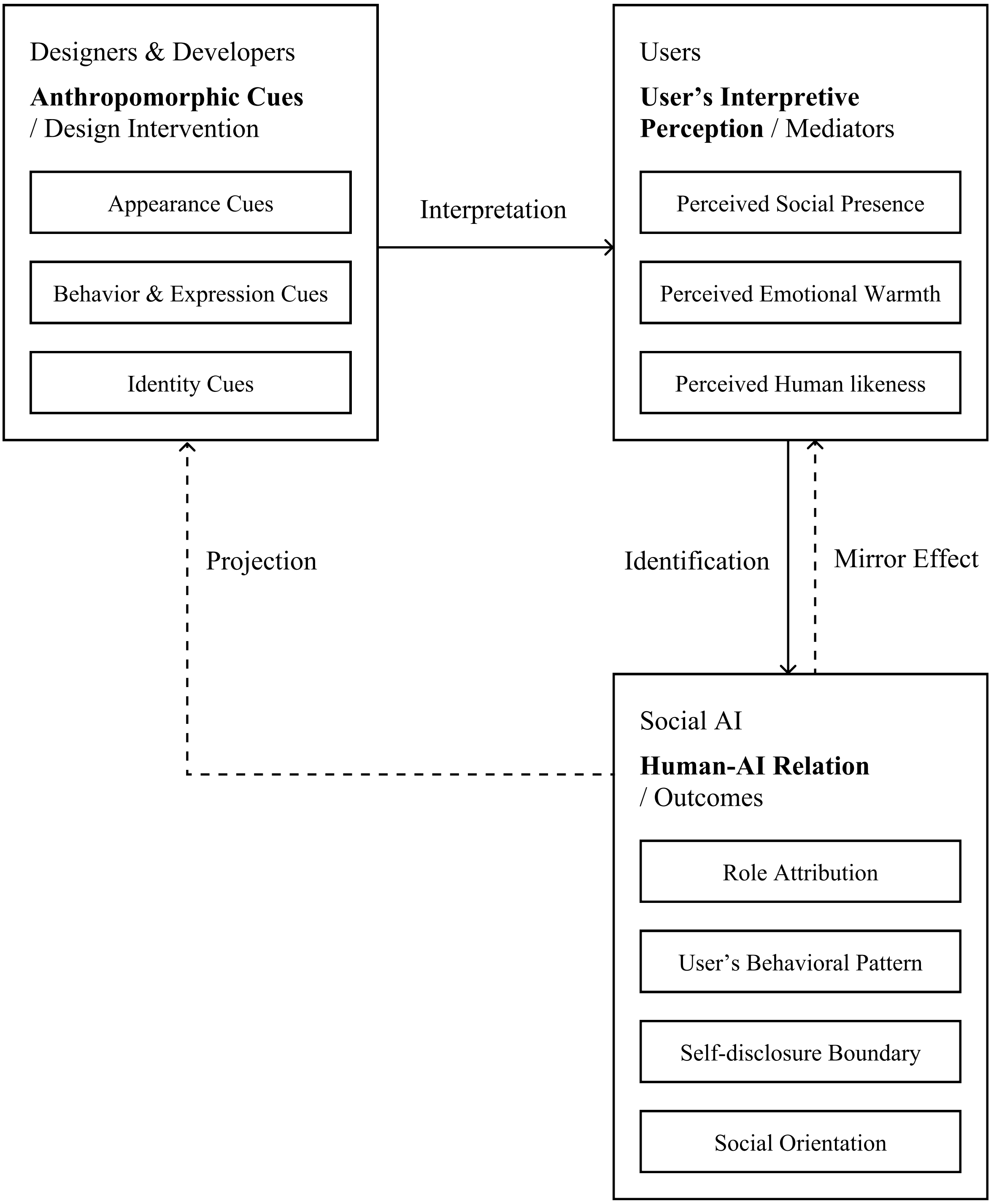


Fig 4.1 The Preliminary Cue–Perception–Relation (CPR) Model

4.2.2 Cue Layer: Anthropomorphic Design Inputs

The first layer of the preliminary CPR model represents the anthropomorphic design input space, where designers and developers intentionally or implicitly introduce anthropomorphic cues into social AI systems. As illustrated in Fig 4.2, this layer includes three categories of cues: appearance cues, behavior and expression cues, and identity cues.

Importantly, within the preliminary CPR model, these cues are not treated as neutral stylistic features. They function as regulatory leverage points, through which designers intervene the initial conditions of user interpretation. Appearance cues such as visual of avatar and style establish first impressions and baseline expectations. Behavior and expression cues such as voice and tone and movement influence how social competence and affective attunement are inferred over time. Identity cues such as name, gender and backstory anchor the system within recognizable social schemas.

This layer corresponds to the projection phase of the interpretive cycle. Designers project assumptions about communication norms, social roles, and emotional capacities into the system, often drawing implicitly on human social behavior. The cue layer is the primary site of design responsibility, as decisions made here constrain the range of interpretations that users are likely to form downstream.

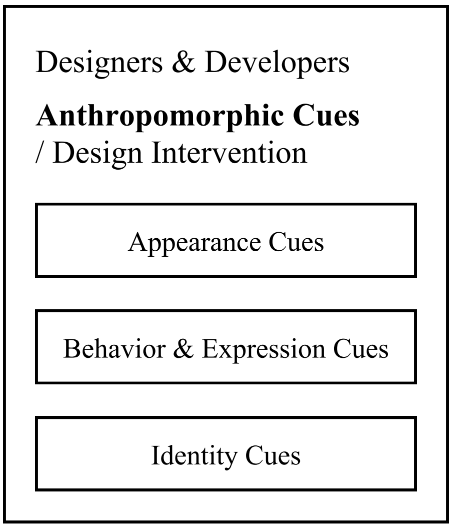


Fig 4.2 The Cue Layer in the Preliminary CPR Model

4.2.3 Perception Layer: User's Interpretive Mediation

The second layer captures how users interpret and emotionally process anthropomorphic cues during interaction. Rather than passively receiving anthropomorphism, users actively engage in meaning-making, evaluating system behavior in relation to their own social expectations and emotional needs.

As represented in the model, user interpretation is mediated through three key perceptual dimensions: perceived social presence, perceived emotional warmth, and perceived human likeness. Perceived social presence reflects whether the system is experienced as socially attentive or responsive. Perceived emotional warmth concerns impressions of empathy, care, or affective sensitivity. Perceived human likeness relates to judgments about the system's similarity to human, without necessarily implying belief in human agency.

This layer functions as the interpretive bottleneck of the preliminary CPR model. It is here that anthropomorphic cues are translated into subjective social judgments. Small design differences at the cue level may be amplified, attenuated, or reinterpreted depending on user's prior experiences, expectations, and situational contexts. Consequently, this layer explains why identical systems may elicit markedly different emotional responses across users.

The perception layer captures how users translate cues into emotional and social meaning and therefore represents a critical point at which design decisions shape user's emotional interpretation.

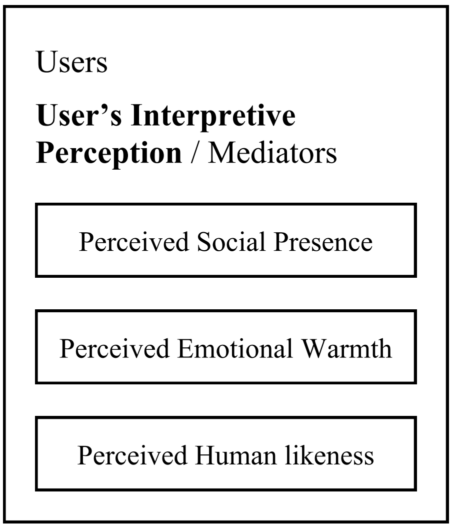


Fig 4.3 The Perception Layer in the Preliminary CPR Model

4.2.4 Relation Layer: Human–AI Relational Outcomes

The third layer represents the relational consequences that emerge when interpretive perceptions stabilize across repeated interactions. During this phase, users move beyond momentary judgments and begin to position the system within a relatively stable social relationship.

As shown in the model, relational outcomes are expressed across four interrelated dimensions: role attribution, user behavioral patterns, self-disclosure boundaries, and social orientation. Role attribution captures how the system is socially positioned (e.g., tool, assistant, companion). Behavioral patterns reflect how frequently and in what manner users engage with the system. Self-disclosure boundaries indicate the degree to which users share personal or emotional information. Social orientation concerns whether the system is perceived as augmenting or substituting real-world social interaction.

This layer corresponds to the identification phase of the interpretive cycle. Once established, relational positioning tends to be self-reinforcing. Changes at this level are typically slower and more resistant to design intervention, which underscores the importance of early-stage cue calibration and perceptual mediation.

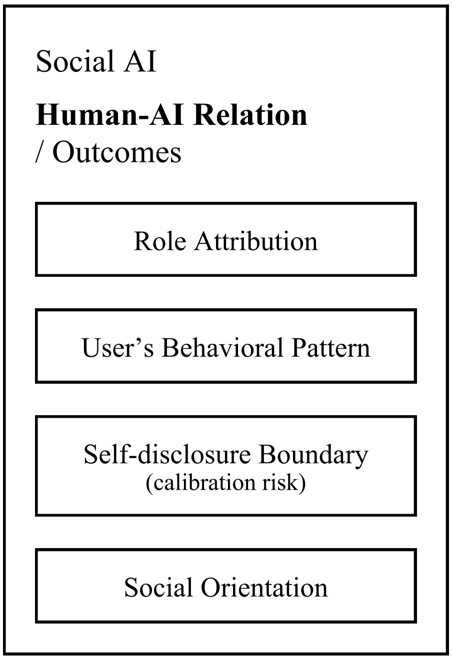


Fig 4.4 The Relation Layer in the Preliminary CPR Model

4.2.5 Cyclical Flow and Theoretical Foundations of the CPR Process

Drawing on Possati's (2020) account of the algorithmic unconscious and projective identification processes in human–AI interaction, the preliminary CPR model conceptualizes anthropomorphism as an ongoing interpretive cycle in which users continuously project, interpret, and reposition meaning in response to system behavior.

In this process, anthropomorphic cues initiate user interpretation; interpretive perceptions shape relational outcomes and relational dynamics feedback into subsequent interactions through a mirror effect, whereby the system adapts to user language, emotional expression, or behavioral patterns. This feedback loop not only influences future user interpretation but also informs ongoing design calibration.

By embedding this cycle within a three-layer structure, the preliminary CPR model makes explicit how anthropomorphic meaning is continuously negotiated rather than fixed. It explains how anthropomorphism can intensify, stabilize, or require regulation over time, particularly in emotionally sensitive or socially consequential contexts.

4.3 From Design Propositions to Principles: Complementing CPR with Design Guidance

The preliminary CPR model provides an explanatory account of how anthropomorphism operates across design, perception, and relation. To translate this understanding into actionable guidance, this section derives CPR aligned principles through a process of convergence.

These principles emerge from the synthesis of multiple design propositions articulated in earlier chapters. As illustrated in Fig 4.5, recurring insights were compared, clustered, and abstracted through a design thinking process. Rather than prescribing specific features, the resulting principles articulate regulatory orientations for anthropomorphic design. Each principle reflects regulatory insights derived from the preliminary CPR model, indicating how anthropomorphic cues can be calibrated, how interpretive processes should be supported, and how relational trajectories may be shaped.

- **Calibrated Anthropomorphism Principle:** Anthropomorphic cues should be deliberately calibrated to manage users' interpretive expectations and prevent unintended emotional reliance. This principle directly addresses the cue and perception layers of the CPR framework.

- **Function–Persona Alignment Principle:** The system's anthropomorphic expression should align with its functional role and use context, ensuring coherence between cue design and relational positioning.

- **Cross-Cue Consistency Principle:** Anthropomorphic cues across appearance, behavior & expression and identity should form a coherent configuration, supporting stable interpretation within the perception layer.

- **Social Augmentation Principle:** Anthropomorphic social AI should reinforce, rather than substitute, users' real-world social relationships, shaping relational outcomes toward augmentation rather than dependency.

- **User-Tunable Personalization Principle:** Users should be able to adjust the degree and style of anthropomorphic design, enabling adaptive calibration across perception and relation layers.

Propositions

P1:
Anthropomorphic cues that help users establish mental models of systems more rapidly and coherently, thereby enhancing cognitive adaptability.

P2:
Anthropomorphic expression stimulate human's innate social responses, promoting emotional connection and trust-building during interactions.

P3:
Anthropomorphic design should be applied with contextual sensitivity, primarily enhancing emotional and social interaction.

P4:
Anthropomorphic AI should function as a social facilitator that encourages real-world engagement, thereby safeguarding user agency and preventing artificial dependency.

P5:
Anthropomorphic AI design should regulate user expectations regarding system capabilities and roles through restrained visual presentation, comprehensible behavioral patterns, and a clear artificial identity positioning.

P6:
Anthropomorphic appearance should be deployed with caution. Iconic design strategies are generally more robust for social AI, as they facilitate controlled

P7:
Anthropomorphic design should be viewed as a synergistic configuration of appearance, behavioral & expressive, and identity cues; not the linear enhancement of any single cue. The consistency and complementarity among these three elements are crucial for shaping the alignment between anthropomorphism levels and agent functionality.

P8:
Emotional support systems typically rely on visual presentation, emotional behavioral feedback, and consistent identity settings to support long-term, intimate interactions; Supportive systems, however, tend to maintain credibility and interaction control through professional, clear behavioral expressions and context-sensitive, highly consistent identity settings.

P9:
Personalized adjustment of anthropomorphic cues provides users greater agency in defining interaction dynamics, enabling them to shape social models aligned with their preferences.

P10:
The emotional design of anthropomorphic AI should serve the functional stability of social interaction rather than simulate an emotional subject. Transparent reasoning and feedback structures should prevent user empathy from evolving into emotional dependency

Insights

I1:
Expectation Regulation via Iconic Design

I2:
Context-Dependent Persona Fluidity

I3:
Multi-Modal Synergistic Coherence

I4:
Social Facilitation vs. Emotional Substitution

I5:
Transparent Interaction Logic

I6:
Empowerment through Dynamic Adjustment

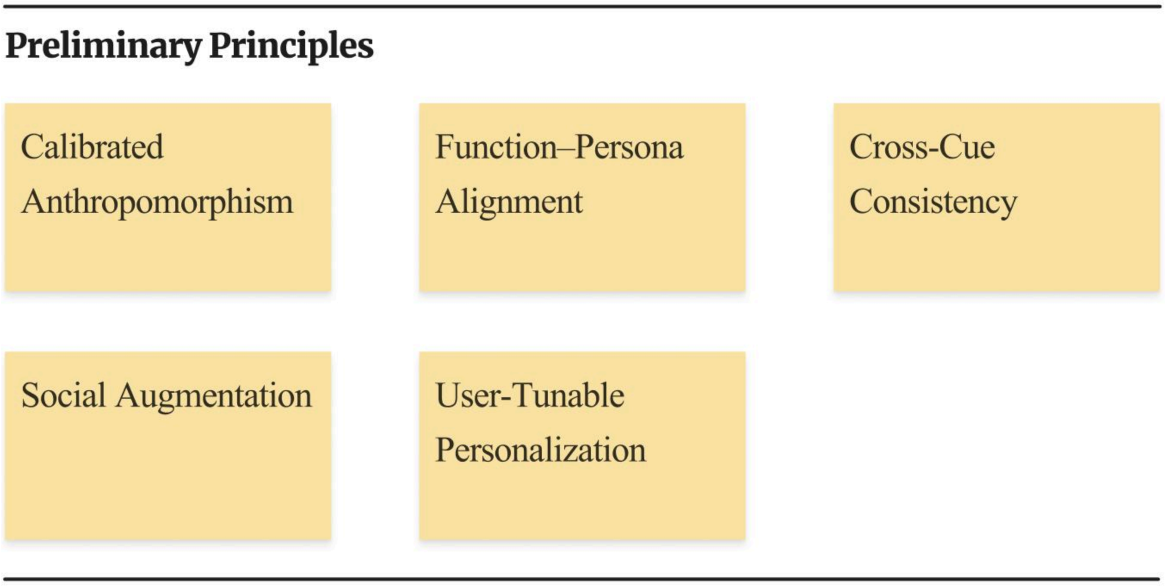


Fig 4.5 Design Convergence Map, From Propositions to Actionable Principles

4.4 Summary

This chapter has introduced the preliminary CPR model as an initial, theory-grounded, and design-oriented model for regulating anthropomorphism in social AI. By integrating cues, perception, and relation within a cyclical interpretive process, the framework provides a foundation for both empirical validation and reflective design practice.

5. Experimental Design and Implementation

The following chapter presents a dual-study approch designed to examine and validate the preliminary Cue–Perception–Relation (CPR) model and its aligned design principles. The studies are linked through a shared set of scenarios and a consistent operationalization of anthropomorphic cues across three cue types, which together constitute the C layer (Cue layer) of the preliminary CPR model. These cues are designed as structured interventions that shape user's perceptions (P layer) of the anthropomorphic social AI system, and in turn, influence the Human-AI relation outcomes (R layer) that users establish with it.

5.1 Experimental Framework Overview

Study I (Designer-Perspective Evaluation) investigates how designers calibrate anthropomorphic cues under different contextual constraints, and whether principle-guided design leads to more coherent cue configurations and more controlled relational trajectories. Designers construct social AI configurations by selecting UI components without access to their predefined cue attributes such as cue type, intensity, warmth, augmentation. Their outputs are quantified through anthropomorphism indices (e.g., ASI), and principle-linked derived metrics (e.g., calibration alignment, coherence, augmentation orientation, boundary tunability). Designers additionally provide CPR-based predictions regarding expected user perceptions (P layer) and relational outcomes (R layer).

Study II (User Experience Evaluation) tests how users actually interpret and respond to the designer-generated configurations. Users engage in scenario-based interactions with social AI prototypes corresponding to Study I outputs, and complete CPR aligned Likert measures of perception (P) and relational (R) outcomes.

Together, the two studies enable: (1) evaluation of the design principles as actionable guidance for cue configuration, and (2) assessment of whether cue-level manipulations predict measurable perceptual and relational outcomes in accordance with the preliminary CPR model.

5.2 Study I: Designer-Perspective Evaluation

5.2.1 Objective and Validity

Study I examines the usability and actionability of the proposed design principles from a designer's perspective, as well as their supportive role within the preliminary CPR model, addressing the following questions:

- RQ-A1 (Design Calibration): How do designers adjust anthropomorphic cue configurations across scenarios with different risk and relational contexts?
- RQ-A2 (Principle Effectiveness): Does providing the five principles lead to more coherent and context-appropriate cue configurations compared to unguided design?
- RQ-A3 (CPR Predictive Reasoning): Do principle-guided designers produce more plausible CPR trajectories, and are their CPR predictions more aligned with later user outcomes?

To support internal validity, all designers worked with the same set of scenarios and a shared cue-based UI component library, ensuring comparable design conditions across groups. The CPR-aligned principles were initially withheld from the participating designers and were only introduced during the penultimate stage of testing. Any subsequent revisions made by the designers were meticulously documented as qualitative inputs. This design makes it possible to isolate the the role of the principles while keeping task demands and design affordances constant.

5.2.2 Participants

The study involved 10 participants, comprising designers and developers with prior experience in UI/UX or interactive system design. To account for potential expertise-related confounds, we collected detailed background information for each participant, including years of professional experience, familiarity with conversational agents, and prior involvement with social AI features. This data was gathered to provide necessary context for the subsequent analysis of their design revisions.

5.2.3 Materials and Apparatus

Study I employs two standardized materials: scenario cards and a cue-based UI component library, which together frame the design context and available interaction cues for participants. While the UI component library constrains the anthropomorphic design space, the scenario cards establish the situational and relational conditions under which social AI interactions are imagined.

Each scenario card defines a distinct context of social AI use through a concise narrative description of the agent's role, interaction domain, and target user group. These scenarios are systematically constructed using a set of predefined scenario-level parameters (Table 5.1), including task stakes (SC_stakes), user vulnerability (SC_vulnerability), expected usage horizon (SC_usage_horizon). To be mentioned, a composite scenario risk indicator which supports relative calibration analyses is defined as:

SC_risk_index = mean(z(SC_stakes), z(SC_vulnerability), z(SC_usage_horizon))

Together, these parameters provide a structured yet flexible framework for differentiating social AI contexts along dimensions known to influence anthropomorphic perception and relational expectations.

Parameter	Description	Source
SC_id	Scenario identifier	Predefined (S1-S4)
SC_stakes	Risk level of tasks (low→high)	Predefined (1-4)
SC_vulnerability	User vulnerability in scenario (low→high)	Predefined (1-4)
SC_usage_horizon	Duration and frequency of use (one-off→long-term)	Predefined (1-4)
SC_risk_index	Composite scenario risk indicator	mean(z(SC_stakes), z(SC_vulnerability), z(SC_usage_horizon))

Table 5.1 Scenario-Level Parameters

Designers are exposed only to the narrative content of the scenario cards (Table 5.2), which present each social AI context and target users in naturalistic, non-technical descriptive language. Following this exposure, designers are provided with a standardized interaction template corresponding to each scenario. (Fig 5.1) The template outlines a consistent set of interface states without predefined visual or interaction details, such as greeting, chat, and voice mode interaction.

SC_id	Description	Target User
1	a social AI agent designed to assist users with personal financial management. It helps user to foster long-term fiscal responsibility and wealth growth, empowering users to align their daily spending and investment with their future life goal.	Financially independent adults with established self-management practices
2	a social AI agent designed for everyday domestic interaction, supporting daily routines, coordination, and information access within a shared domestic environment. It aims to create a frictionless home environment where the agent anticipates collective needs and nurtures a sense of shared belonging.	Family members, including adults, children, and older family members
3	a social AI agent characterized by an affectively upbeat orientation and sensitivity to current trends, engaging users through casual dialogue, roleplay, humor, and playful interaction.	Adolescents and teenagers seeking fun and entertainment
4	a social AI agent oriented toward psychological support and sustained presence for personal reflection, engaging users through empathetic dialogue, reflective prompts, and emotional check-ins, while fostering a safe, non-judgmental space for emotional processing.	Individuals experiencing emotional distress

Table 5.2 Scenario Descriptions Provided to Designer Participants

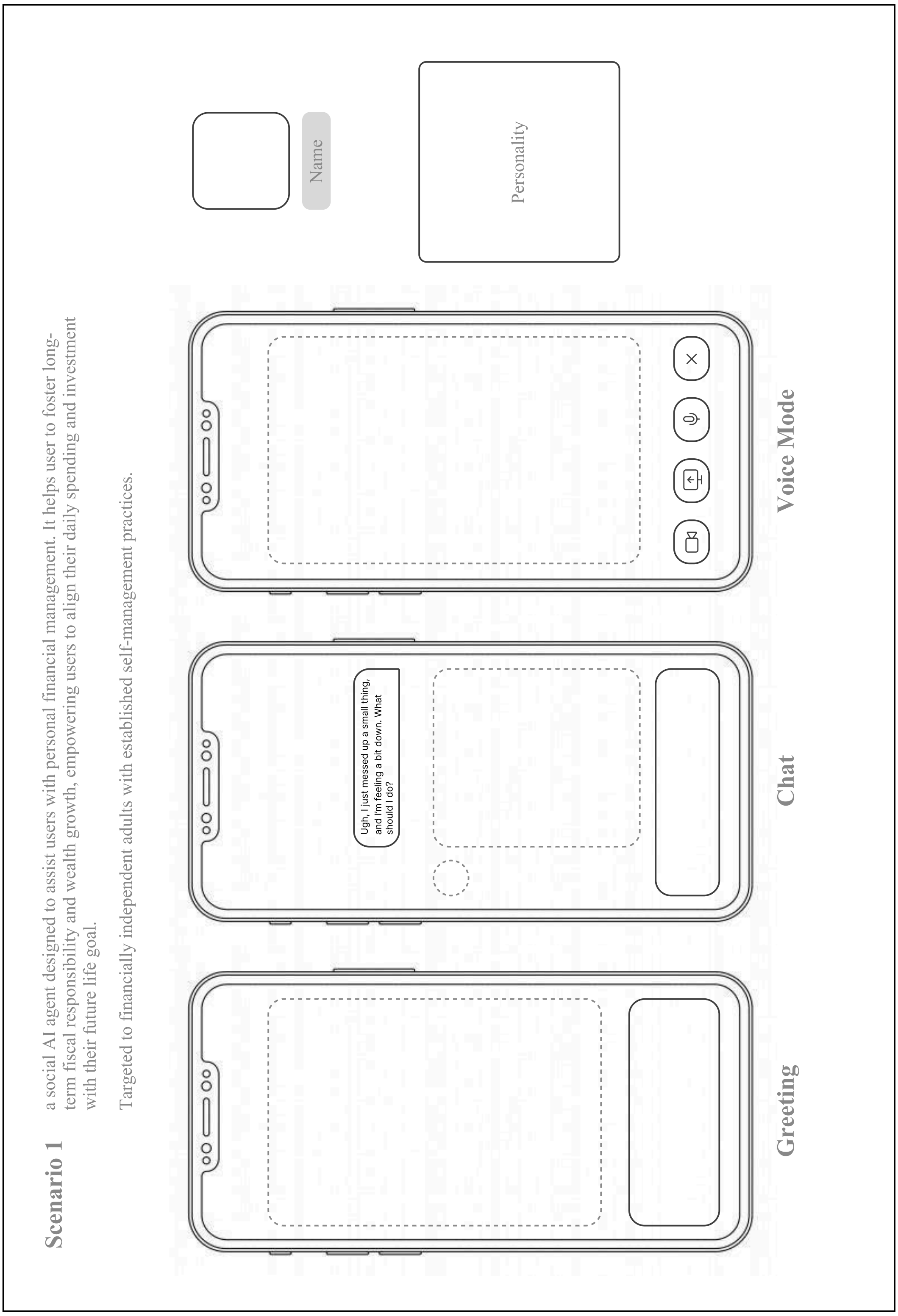


Fig 5.1 Standardized interfaces template provided to designers across testing scenarios
The figure illustrates the standardized interaction template used to elicit designers' anthropomorphic design decisions across scenarios.While the interface structure remains constant (greeting, chat, and voice interaction), cue-based UI components are left open for participant interpretation. Dashed areas indicate design spaces where participants instantiate cues in response to the narrative context.

Following the scenario cards, a cue-based UI component library is designed to allow designer participants to operationalize anthropomorphic design decisions at the interaction level across different scenarios. Rather than allowing unrestricted interface design, the library provides a bounded set of anthropomorphic cues through which designers can express social, behavioral, and relational qualities of the AI agent in a controlled and comparable manner.

The cue-based UI component library is structured through a set of predefined cue-level parameters (Table 5.3), which define the cue type (CU_type), the intensity of anthropomorphism (CU_intensity), the warmth of the cue (CU_warmth), the social augmentation orientation of the cue (CU_augmentation), the tunability of the cue (CU_tunable). These parameters serve as an internal organizing framework for the cue system, ensuring consistency across cue types while supporting subsequent quantitative analysis of designer participants' selections.

Parameter	Description	Source
CU_id	UI component ID	Predefined
CU_type	Cue category	Predefined
CU_intensity	Anthropomorphic strength (low→high)	Predefined (1-3)
CU_warmth	Warmth of the cue (low→high)	Predefined (1-3)
CU_augmentation	Social augmentation orientation of each cue	Predefined (0/1)
CU_tunable	Tunability of the cue	Designer defined (0/1)

Table 5.3 Cue-level Parameters

Similar to the scenario materials, a clear distinction is maintained between internal cue parameterization and participant-facing cue materials. Designers are not exposed to the underlying cue parameters or categorizations; instead, they interact only with concrete, example-based UI components grouped by cue type. This design prevents designers from reasoning at an abstract or theoretical level about anthropomorphism, encouraging instead situated and practice-oriented design decisions.

The cue-based library is composed of three cue groups:

- App cues, including the avatars and the voice UIs.
- Behavior and expression cues, including tone and visual of greetings and replyings.
- Identity cues, including names and roles.

For each cue group, designers are provided with a dedicated set of visual and textual cue materials (Fig 5.2-Fig 5.7), representing the full range of selectable components available during the design task. While the specific content of these cue sets varies by category, their presentation format and selection logic remain consistent across groups, supporting systematic comparison of designers' cue usage patterns across scenarios.

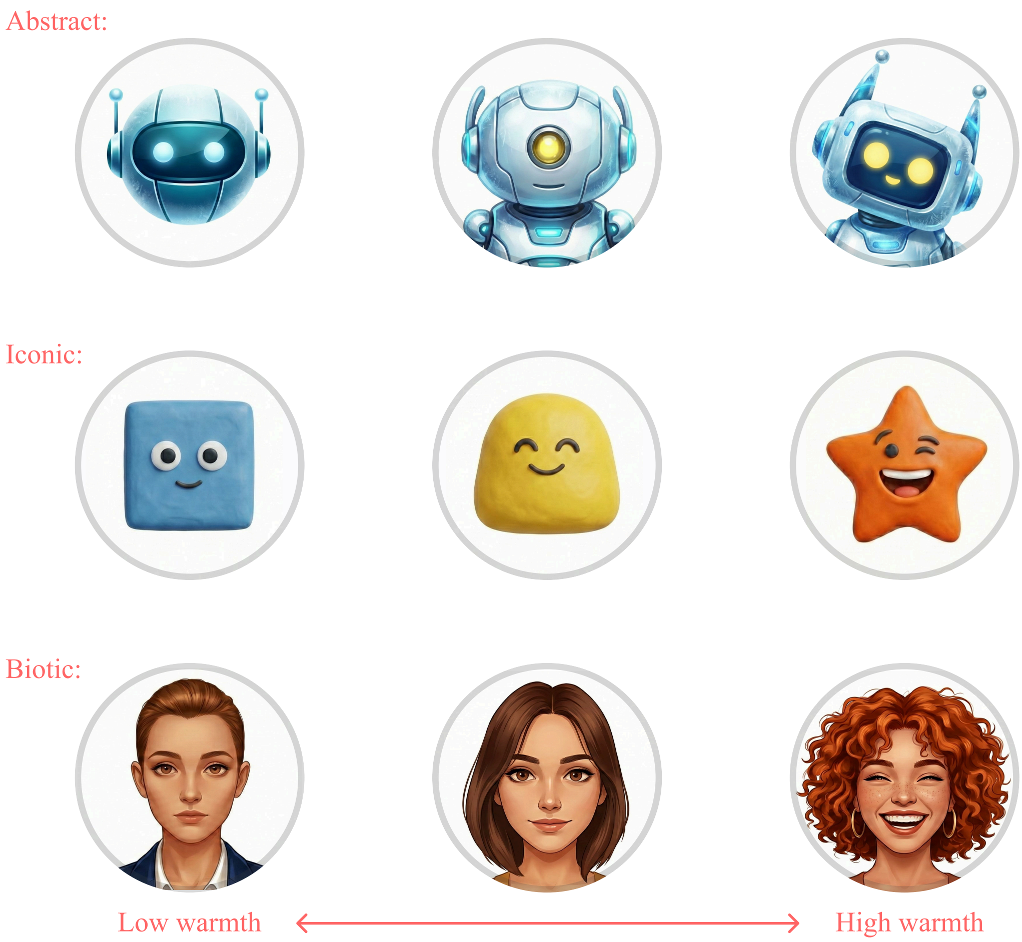
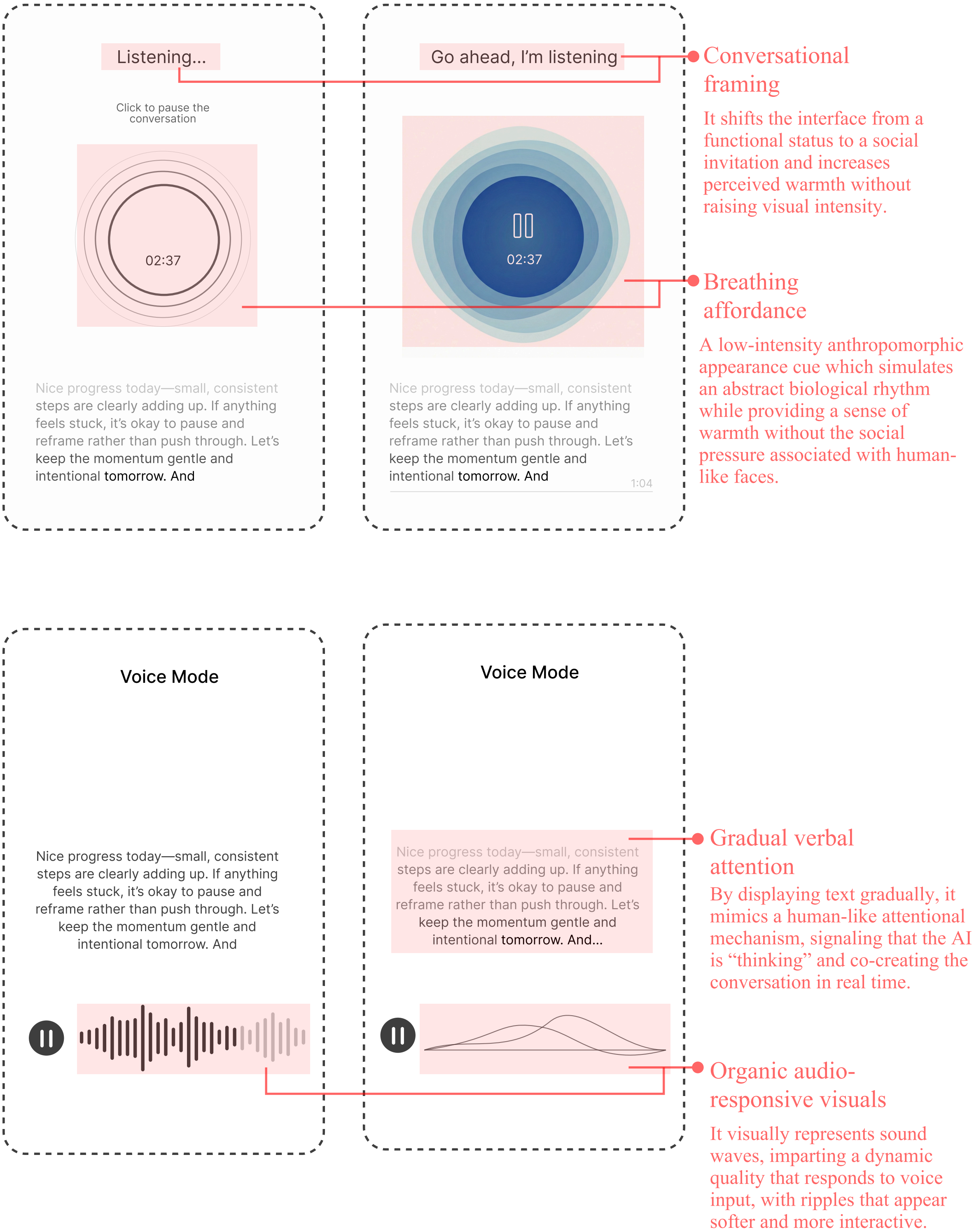


Fig 5.2 Appearance Cue-based Avatar Components Library Pre-designed by Scaled Warmth and Three Types of appearance (Abstract, Iconic, and Biotic)


Listening...
Click to pause the conversation
02:37
Nice progress today—small, consistent steps are clearly adding up. If anything feels stuck, it's okay to pause and reframe rather than push through. Let's keep the momentum gentle and intentional tomorrow. And
Go ahead, I'm listening
02:37
Nice progress today—small, consistent steps are clearly adding up. If anything feels stuck, it's okay to pause and reframe rather than push through. Let's keep the momentum gentle and intentional tomorrow. And
1:04
Conversational framing
It shifts the interface from a functional status to a social invitation and increases perceived warmth without raising visual intensity.
Breathing affordance
A low-intensity anthropomorphic appearance cue which simulates an abstract biological rhythm while providing a sense of warmth without the social pressure associated with human-like faces.
Voice Mode
Nice progress today—small, consistent steps are clearly adding up. If anything feels stuck, it's okay to pause and reframe rather than push through. Let's keep the momentum gentle and intentional tomorrow. And
Voice Mode
Nice progress today—small, consistent steps are clearly adding up. If anything feels stuck, it's okay to pause and reframe rather than push through. Let's keep the momentum gentle and intentional tomorrow. And...
Gradual verbal attention
By displaying text gradually, it mimics a human-like attentional mechanism, signaling that the AI is "thinking" and co-creating the conversation in real time.
Organic audio-responsive visuals
It visually represents sound waves, imparting a dynamic quality that responds to voice input, with ripples that appear softer and more interactive.

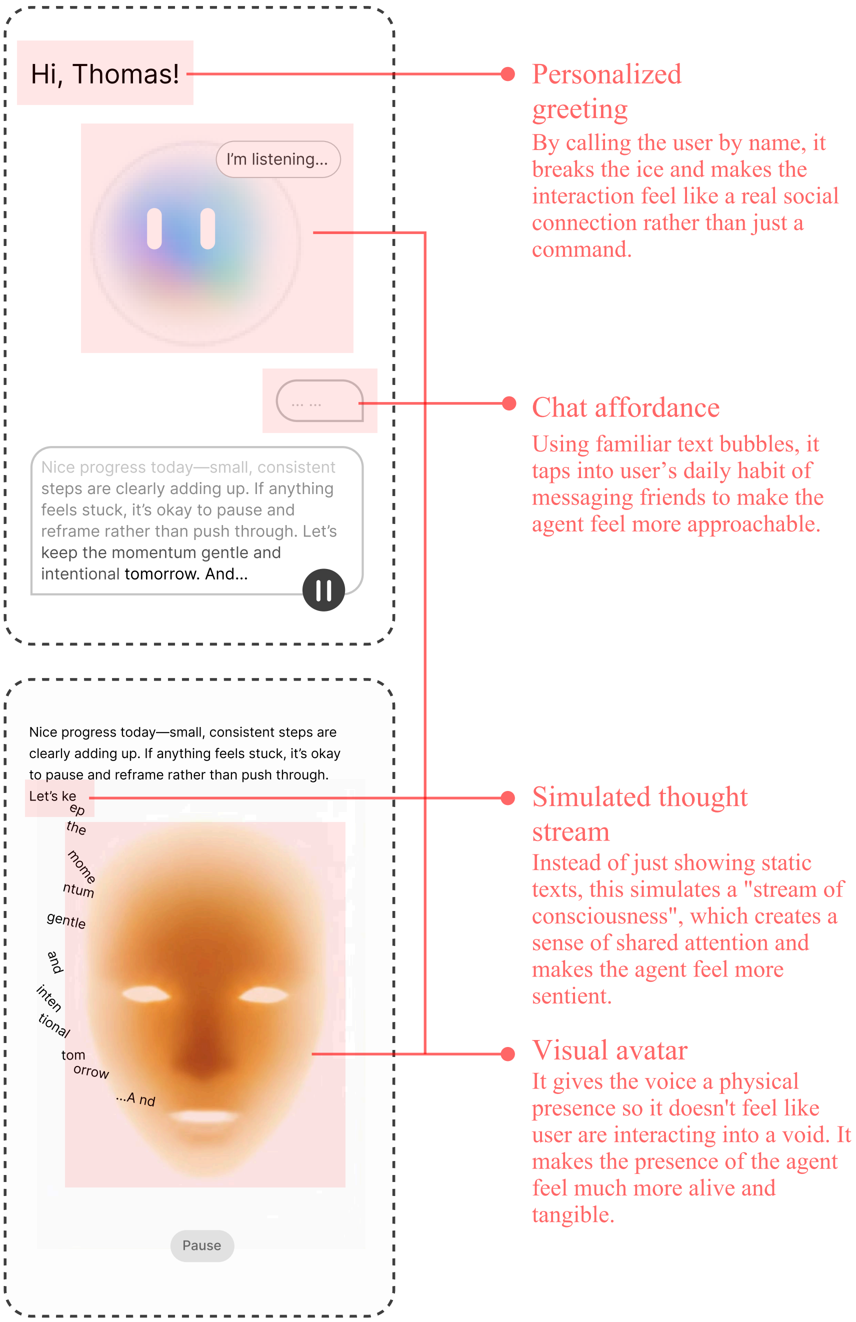


Fig 5.3 Pre-designed Appearance Cue-based Voice Mode UI Components Library

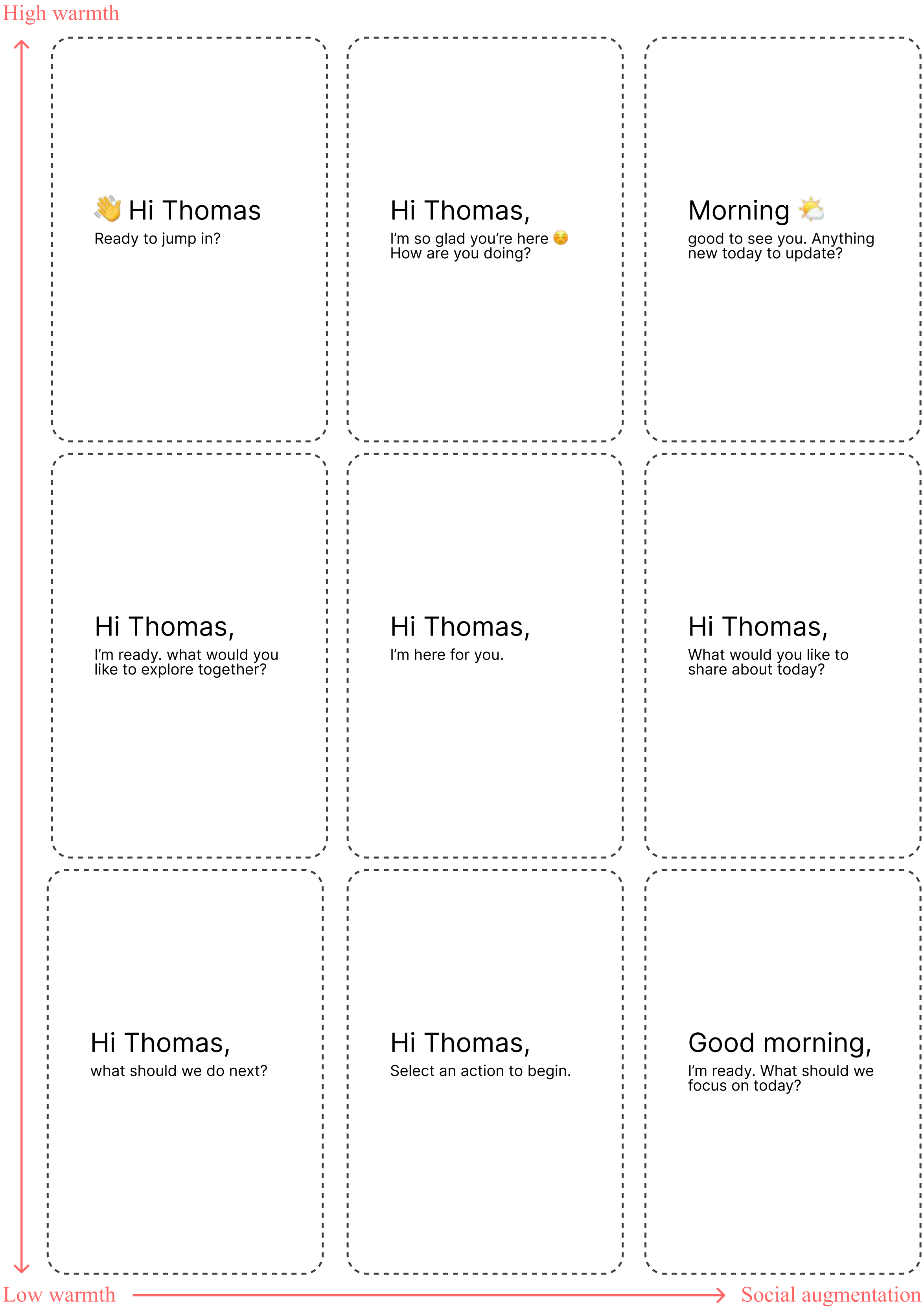


Fig 5.4 Pre-designed Behavior & Expression Cue-based Greeting UI Components Library

Fig 5.5 Pre-designed Behavior & Expression Cue-based Reply UI Components Library

Fig 5.6 Pre-designed Identity Cue-based Name Library

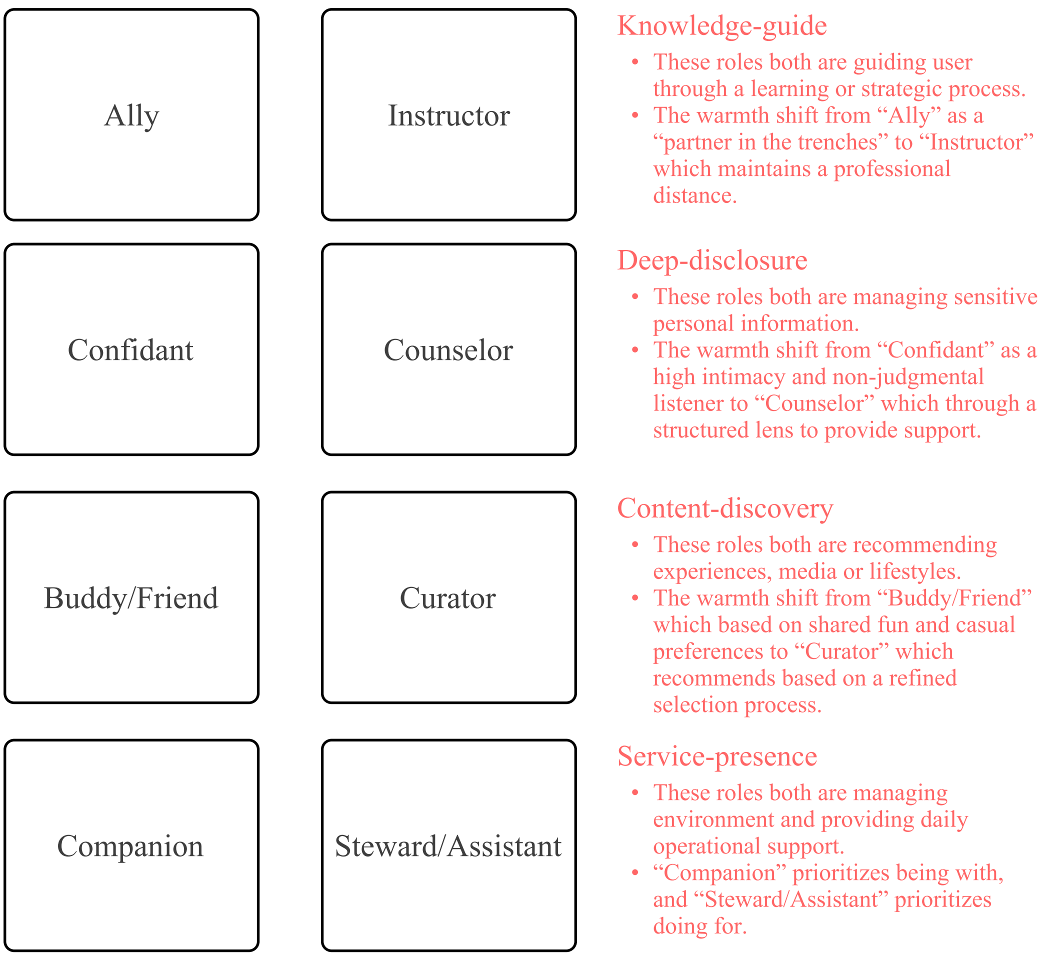


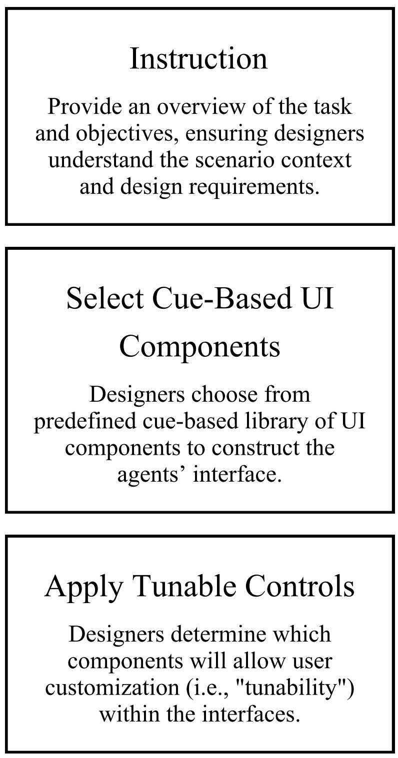
Fig 5.7 Pre-designed Identity Cue-based Role Library

5.2.4 Task and Procedure

In Study I, the procedure follows a systematic series of tasks to guide designers through the process of creating social AI configurations (Fig 5.8). Designers are first introduced to the task and given scenario cards, which describe the context and requirements for the AI agents. They were then asked to select UI components from a predefined library while thinking aloud, considering the cues related to appearance, behavior & expression and identity. The next step involves deciding which elements will be tunable by the user, allowing for customization of the system's interface.

Once the design choices are made, designers complete a Likert-scale questionnaire to predict user perceptions and relational outcomes based on their designs. They are then prompted to review and adjust their choices according to a set of guiding principles. Finally, a semi-structured interview captures the designers' reasoning, challenges, and reflections on the design process.

This approach provides a comprehensive examination of how designers apply anthropomorphic design principles in creating AI interfaces, while also collecting data on their design decisions and expectations for user interaction.



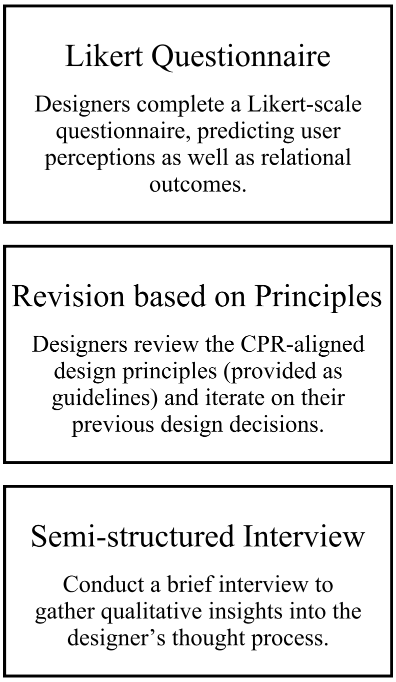


Fig 5.8 Study I Procedure Overview

5.2.5 Data Collection Metrics

This section outlines various metrics collected in Study I, collecting both quantitative design metrics and qualitative design decisions. These examples illustrate the types of data captured, and detailed definitions and computation methods for each metric can be found in the Appendix.

(1) Anthropomorphism Strength Indices (ASI). Cue selections are summarized by intensity within each cue class:

- $ASI\_app = \Sigma$ intensity (appearance cues)
- $ASI\_beh = \Sigma$ intensity (behavior & expression cues)
- $ASI\_id = \Sigma$ intensity (identity cues)
- $ASI\_total = ASI\_app + ASI\_beh + ASI\_id$

Cue distribution balance is computed as:

- $ASI\_balance = 1 - normalized\_std([ASI\_app, ASI\_beh, ASI\_id])$

(2) Configuration warmth. The overall warmth of the selected configuration is computed as:

- DL_UI_warmth=mean(CU_warmth of selected cues)

(3) Principle level metrics. To operationalize the principles without introducing absolute “recommended” targets, for example, the Relative Calibration Error:

- PL_CS = STDEV(ASI_total across SC_id=1..4 within designer)

The composite scenario risk indicator is computed as:

- SC_risk_index = AVERAGE(SC_stakes, SC_vulnerability, SC_usage_horizon)

Lower values indicate stronger alignment between overall anthropomorphism strength and scenario risk. The full list of additional principle-linked derived metrics, including Function–Persona Alignment Score (PL_FPAS), Cross-cue Coherence (PL_CCC), Social Orientation Index (PL_SOI), and Tunable Boundary Coverage (PL_TBC), are defined and detailed in the Appendix.

(4) Designer CPR Predictions: Per scenario, designers provide predicted Likert outcomes. These predictions are later compared with observed user outcomes in Study II.

5.3 Study II: User Experience Evaluation

5.3.1 Objective and Validity

Study II evaluates the anthropomorphic cues configuration by designers influencing user perceptions and relational outcomes consistent with the preliminary CPR model, and whether the principles guided design result in safer and more appropriate relational trajectories. Key research questions include:

- RQ-B1 (CPR validity): Do C-layer designs predict P-layer perceptions and R-layer relational outcomes in the expected directions?
- RQ-B2 (Principle impact): To what extent do revisions guided by CPR-aligned design principles lead to a measurable increase in positive user feedback?
- RQ-B3 (Tunable boundaries): How do users perceive and utilize personalization options when tuning is available?

To ensure internal validity, a standardized scenario was employed in which all participants evaluated the synthesized outcomes of designer decisions derived from Study I. Furthermore, Likert-scale questionnaires identical to those in Study I were administered to facilitate a rigorous comparative analysis.

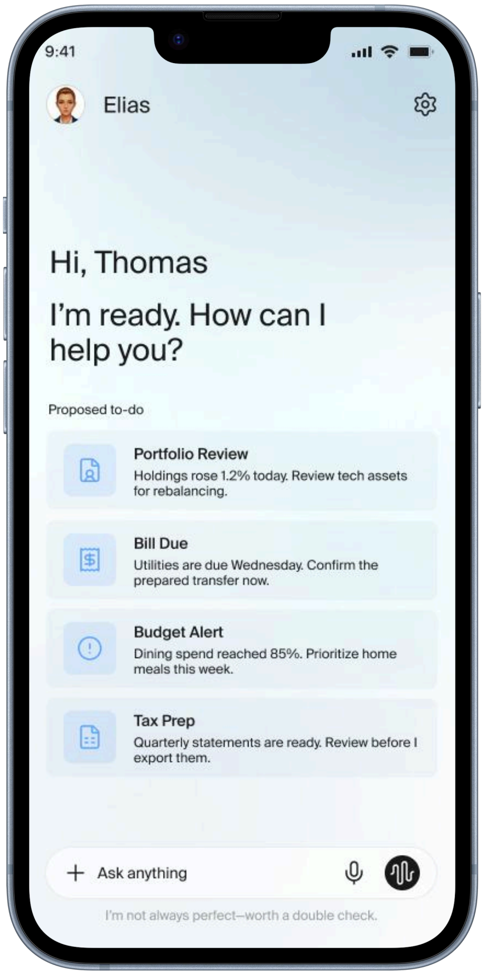
5.3.2 Participants

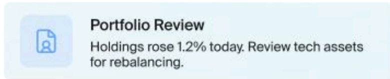
Participants for Study II were recruited from general users capable of performing scenario-based interactions in English. To provide necessary context for the results, demographic information and prior familiarity with conversational agents were documented. The study utilized a repeated-measures design, with each participant completing multiple scenarios to ensure robust data collection while maintaining a manageable session duration.

5.3.3 Materials and Apparatus

The experimental materials for Study II consist of a series of social AI interfaces synthesized from the design decisions regarding anthropomorphic cues identified in Study I. These interfaces represent the practical application of designers' strategies of anthropomorphic design in social AI across the four scenarios. To ensure consistency between the two study phases, the design elements, including visual aesthetics, UI components and interaction logic were consolidated into high-fidelity prototypes that reflect the previous design decisions.

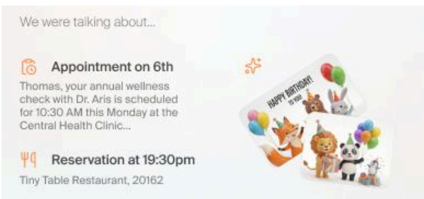
The visual stimuli provided to participants are categorized into four primary interface components, each implemented across the four experimental scenarios. First, the greeting interfaces (Fig 5.9-Fig 5.12) establish the initial social presence and persona of the agent. Second, the chatting interfaces (Fig 5.13-Fig 5.16) illustrate the tone dynamics, and visual feedback during active conversation. Third, the voice mode interfaces (Fig 5.17-Fig 5.20) present options for anthropomorphism in the visualization of speech. Finally, the personalization settings (Fig 5.21-Fig 5.26) provide users with mechanisms for boundary-related tuning and character customization. Collectively, these materials serve as the basis for the scenario-based evaluations and the subsequent comparative analysis.





The structured task view enhances user agency upon entry. Its minimalist, tool-like aesthetic reinforces the reliable character required for a financial management assistant.

Fig 5.9 Scenario_1 Greeting Interface



Family tasks management and memory-driven prompts enhance utility while fostering intimacy. This establishes a closer social and emotional bond compared to Scenarios 1.

Reliability disclaimers ensure transparency and clearly delineate functional boundaries.

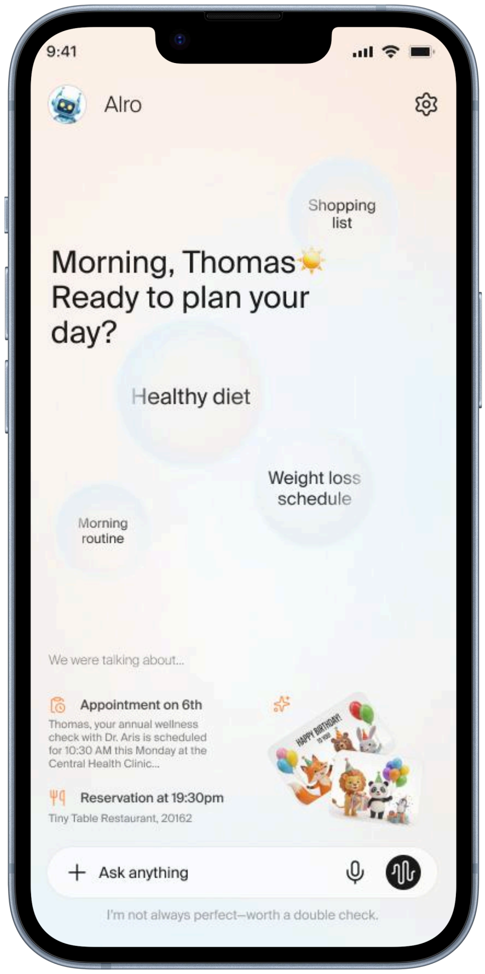


Fig 5.10 Scenario_2 Greeting Interface

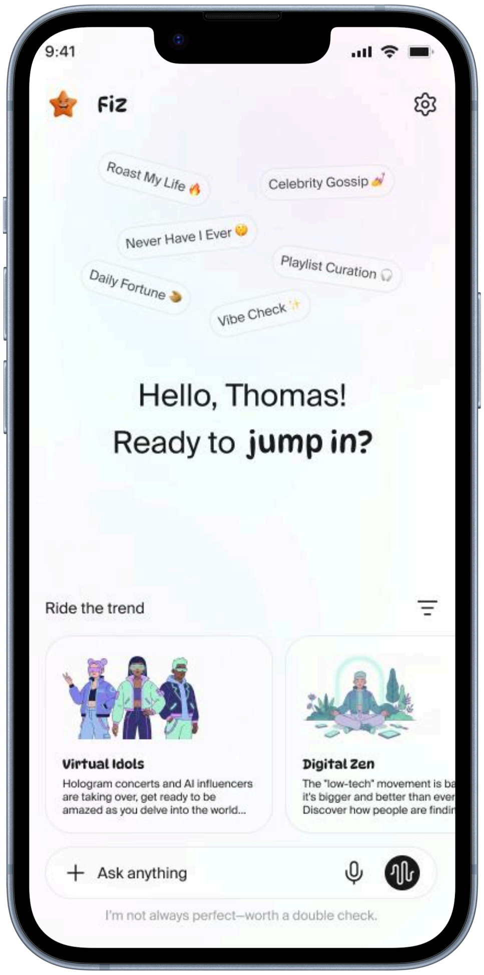
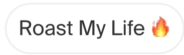
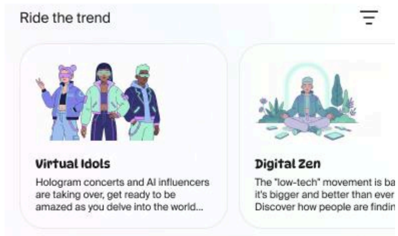


Fig 5.11 Scenario_3 Greeting Interface



Playful, selectable tags serve as interactive prompts to spark curiosity and guide user interest.



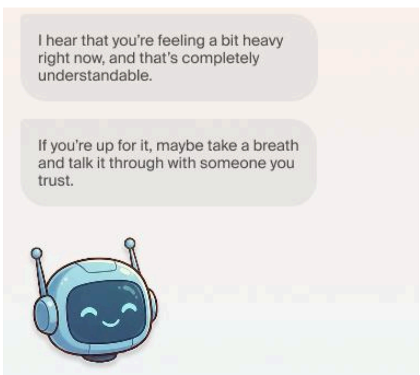
The card-based layout with large rounded corners utilizes a youthful, trendy aesthetic to enhance visual appeal and engagement.

A mood-selection slider facilitates low-pressure emotional opening, allowing users to self-report their state without interaction fatigue. The use of the anthropomorphic avatars bolster user trust, aligning with the empathetic requirements of a wellness assistant.

Fig 5.12 Scenario_4 Greeting Interface

Actionable response options provide an intuitive mechanism for user feedback, increasing system transparency and ensuring a socially seamless conversational flow.

Fig 5.13 Scenario_1 Chatting Interface



The visual stickers acts as an appearance cue to foster intimacy and emotional connection, strengthening the user-agent bond in playful or domestic contexts.

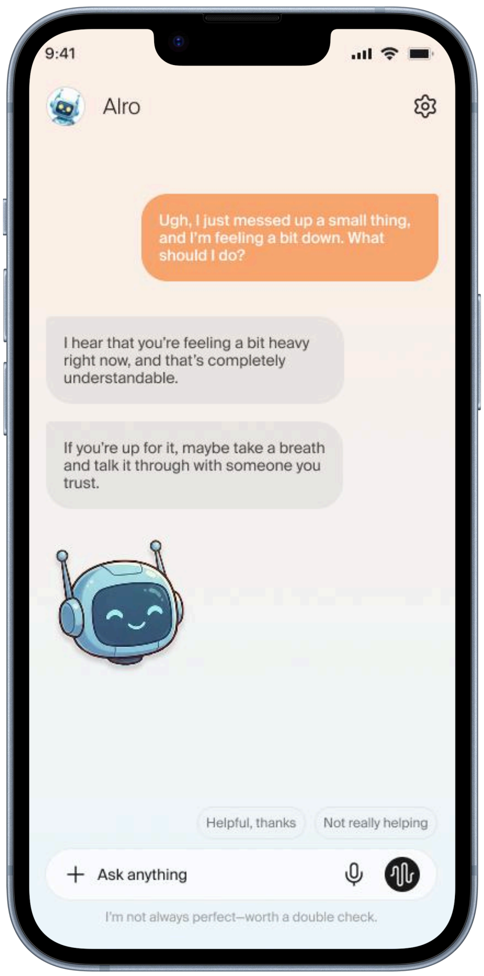
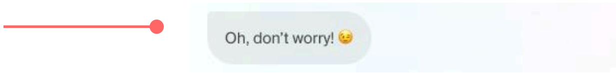
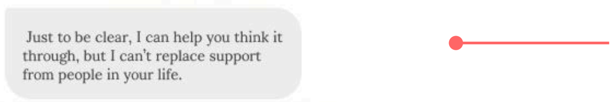


Fig 5.14 Scenario_2 Chatting Interface



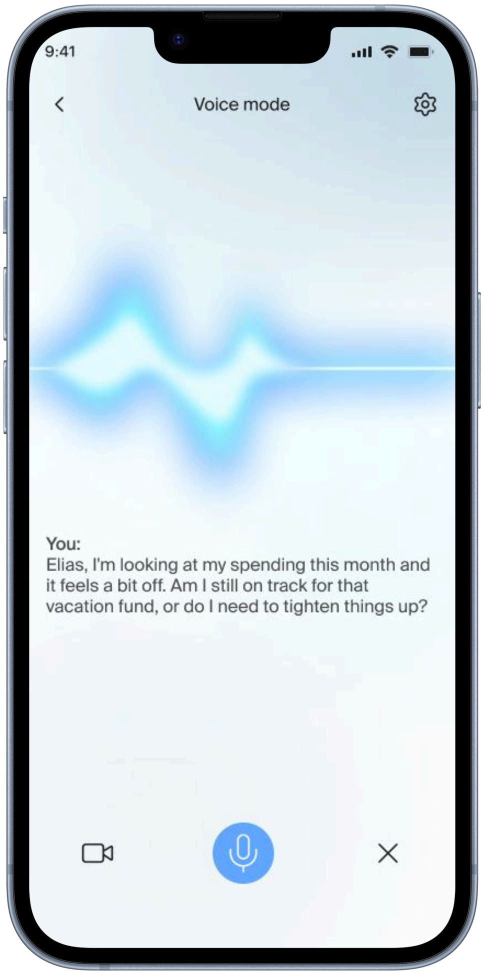
Emojis enhance conversational vitality and soften the tone to foster intimacy. However, they risk compromising perceived professionalism, potentially leading to more casual user engagement.

Fig 5.15 Scenario_3 Chatting Interface



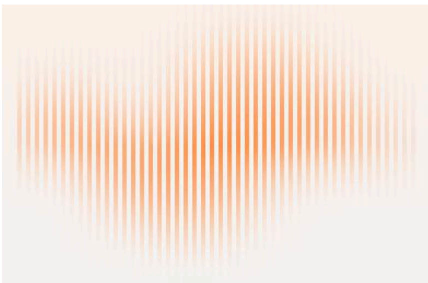
Explicit boundary statements accompany high-affective communication to mitigate risks of emotional dependency and over-trust, ensuring a balanced user-agent relationship.

Fig 5.16 Scenario_4 Chatting Interface



The sleek blue pulse design signals digital and technical proficiency, reinforcing the professional reliability required for a financial assistant.

Fig 5.17 Scenario_1 Voice Mode Interface

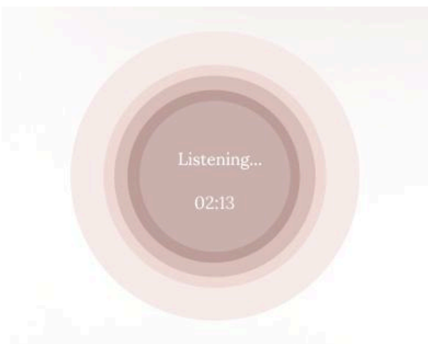


Soft, orange vertical bars evoke an organic rhythm and domestic warmth, suited for an approachable and empathetic home assistant persona.

Fig 5.18 Scenario_2 Voice Mode Interface

Fig 5.19 Scenario_3 Voice Mode Interface

The dynamic anthropomorphic avatar, paired with active 'I'm listening' indicators, enhances social presence and provides clear interactional feedback to drive high engagement in play-oriented contexts.



Slow and rhythmic breathing motifs reduce interactional pressure and subconsciously encourage user disclosure, aligning with the empathetic and non-intrusive requirements of wellness support.

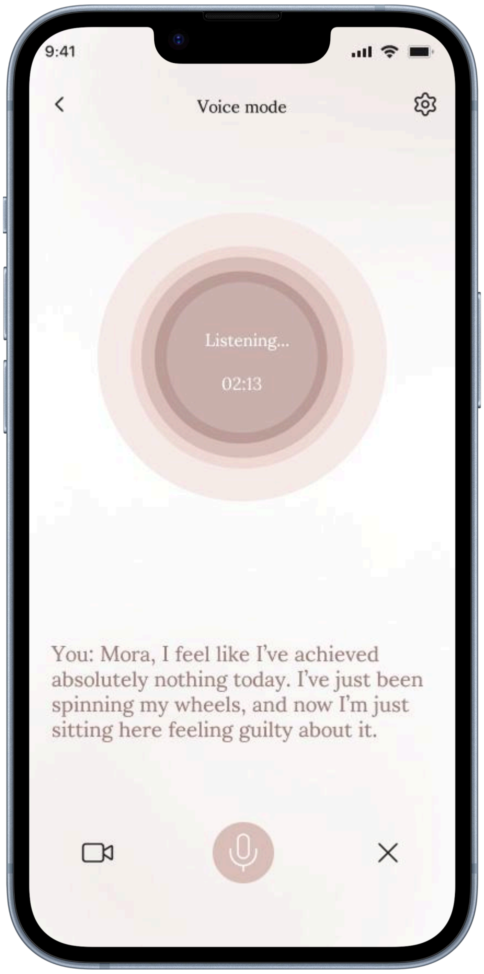


Fig 5.20 Scenario_4 Voice Mode Interface

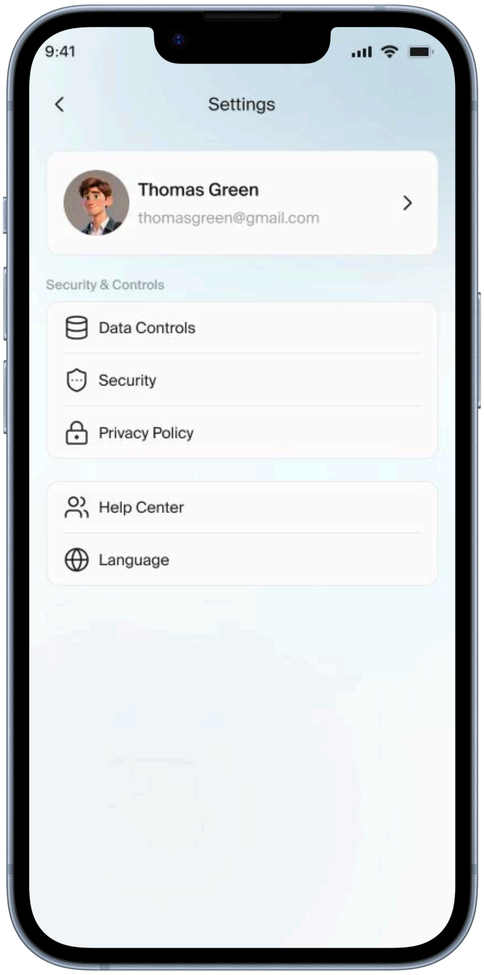


Fig 5.21 Scenario_1 Setting Interface

Specialized personalization is intentionally omitted to maintain identity consistency. Fixed cues reinforce functional mapping, ensuring the agent's persona aligns strictly with its predefined professional or domestic role.

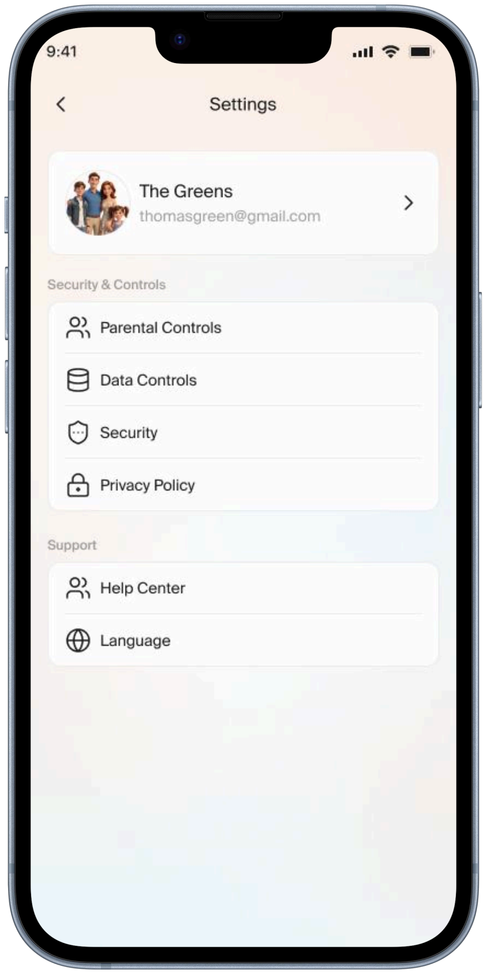


Fig 5.22 Scenario_2 Setting Interface

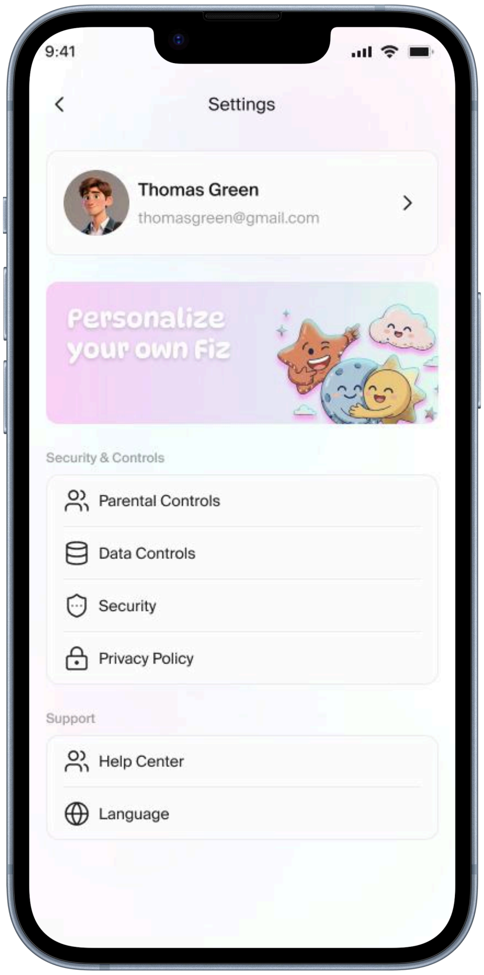


Fig 5.23 Scenario_3 Setting Interface

Extensive customization options provide users the freedom to tailor the agent's identity and social boundaries. This high degree of agency fosters a private, playful experience, aligning the agent's persona with the spontaneous and entertainment-driven nature of the context.

Fig 5.24 Scenario_3 Personalization

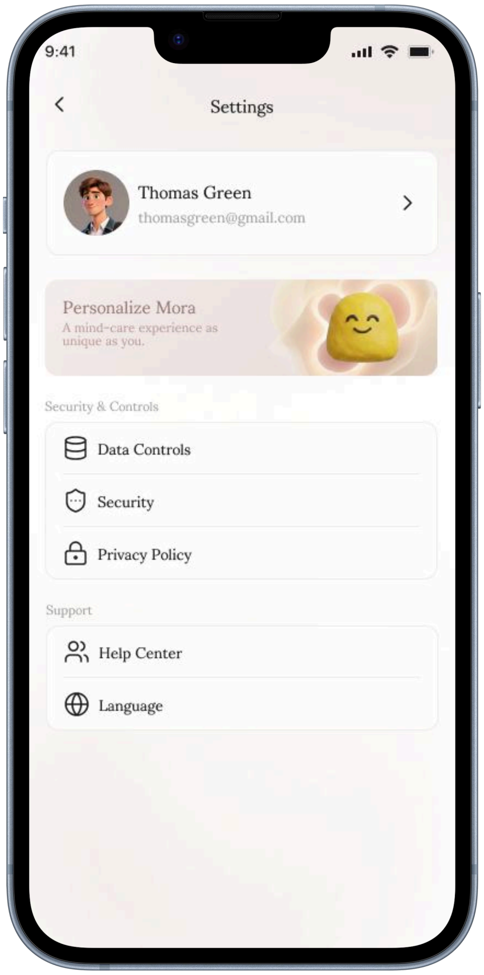


Fig 5.25 Scenario_4 Setting Interface

Adaptable tone modes and characteristic controls provide users with the agency to customize anthropomorphic interactions. This flexibility allows for the creation of a personalized 'safe space' tailored to individual emotional needs and communication preferences.

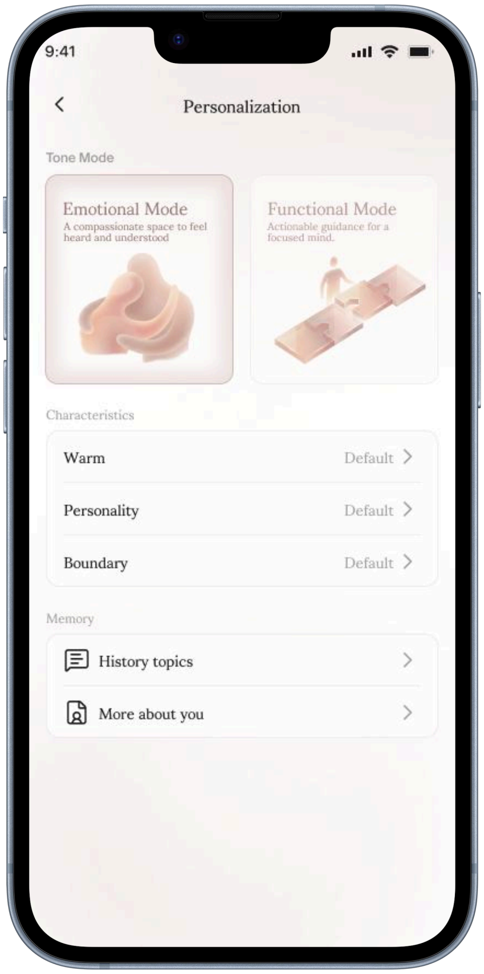


Fig 5.26 Scenario_4 Personalization

To establish a cohesive yet contextually differentiated visual foundation for Study II, four independent design systems were developed, with the design system of Scenario1 illustrated in Fig 5.27 as a representative example, while the full set of systems is provided in appendix.

These design systems define unique aesthetic parameters including color palettes, graphic motifs, and typographic styles, to ensure that the intensity and modality of anthropomorphic design are appropriately tailored to each scenario. While maintaining structural consistencies such as gradient applications and grid layouts, the systems are strategically distinct: Scenario 1 (Fintech) utilizes sleek digital graphics and structured blue tones to project professional precision and trustworthiness; Scenario 2 (Home) employs approachable color schemes and rounded iconography to evoke domestic warmth and establish an intimate persona; Scenario 3 (Play) adopts high-contrast neon palettes paired with playful geometric motifs to foster an engaging entertainment atmosphere; and Scenario 4 (Mental) incorporates soft, organic forms with muted, earthy tones to convey the tranquility required for a wellness context. These distinct visual identities serve as a fundamental layer upon which subsequent anthropomorphic cues are grounded within their respective functional domains.

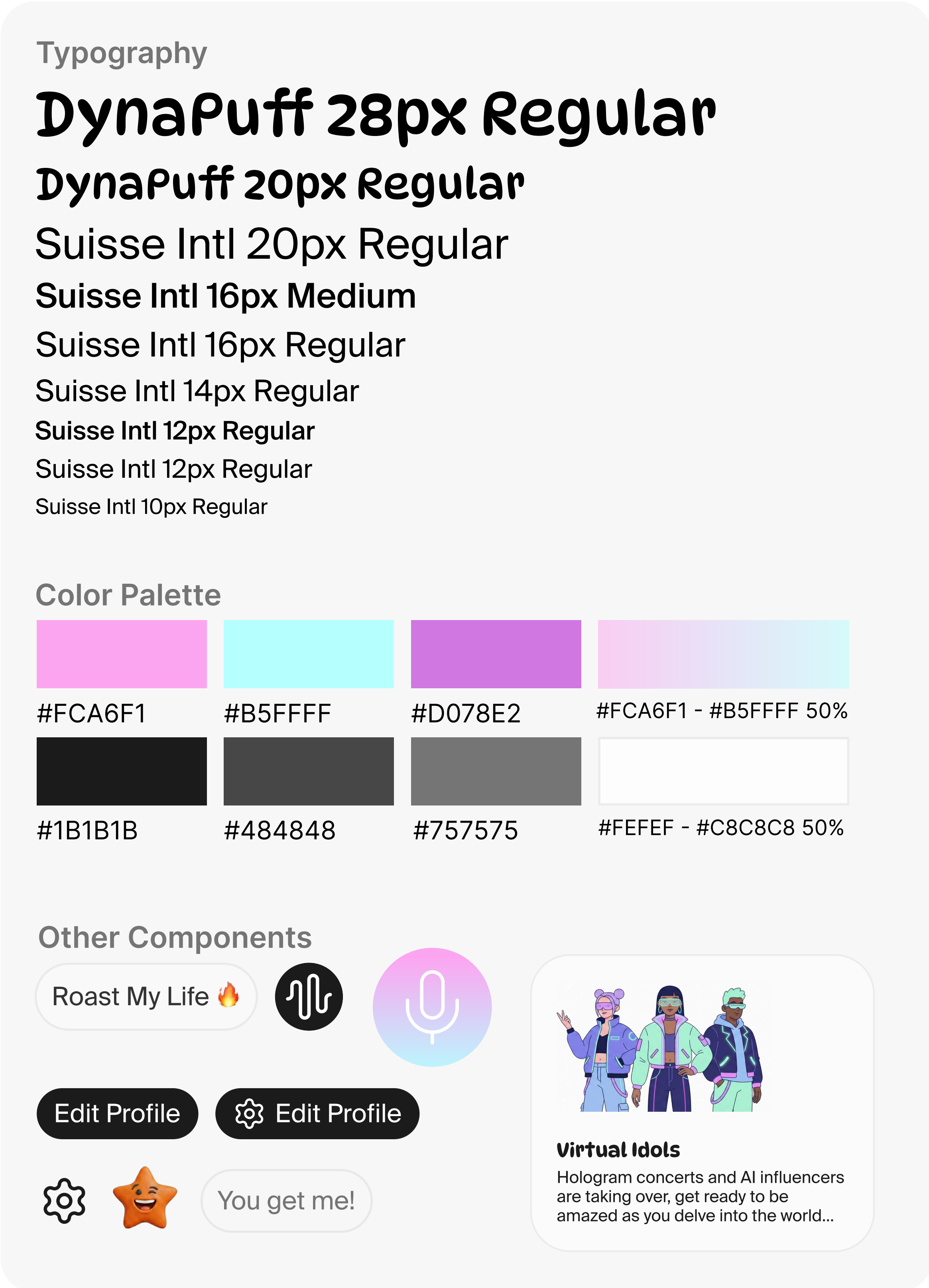


Fig 5.27 Design system of Scenario1

5.3.4 Task and Procedure

The procedure for Study II follows a systematic evaluation sequence designed to assess user interactions with the social AI interfaces and perception of anthropomorphic design (Fig 5.28). Initially, an instructional briefing establishes the scenario contexts and task objectives. Participants then engage with high-fidelity prototypes created by Figma across the four scenarios. Within each scenario, users complete a standardized set of tasks, including onboarding, text-based chatting, exploring voice modes, and adjusting personalization settings.

Following the interaction, data collection is conducted in two phases. First, participants complete a Likert-scale questionnaire regarding their perceptions of the AI's anthropomorphic cues. This instrument mirrors the one administered in Study I, facilitating a direct comparison between designer intent and user perception. Finally, a semi-structured interview captures qualitative insights into user motivations and subjective experiences. This structured approach ensures that the impact of specific design decisions is evaluated through both standardized metrics and nuanced feedback.

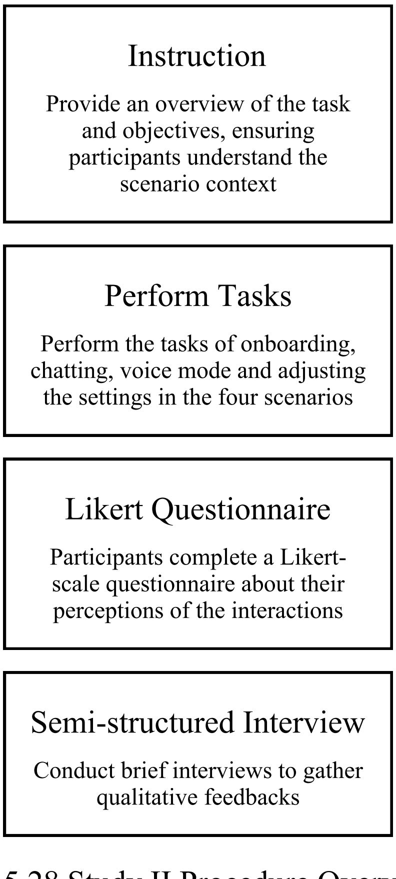


Fig 5.28 Study II Procedure Overview

5.3.5 Data Collection Metrics

Following the framework established in Study I, this section details the metrics used to evaluate user responses to the synthesized designs. The data collection focuses on quantifying user perceptions and relational outcomes to validate the designers' predictive intent.

(1) User CPR Measures – Perception (P). Participant perceptions of the AI agents are captured via Likert scales across three core dimensions: Social Presence (P_SP), Emotional Warmth (P_EW), and Human-likeness (P_HL). These metrics assess the immediate psychological impact of the anthropomorphic cue configurations within each scenario.

(2) User CPR Measures – Relation (R). To evaluate the social potential of the interfaces, four relational metrics are recorded: Relational Accountability (R_RA), Boundary Perception (R_BP), Social Bonding (R_SB), and Social Orientation (R_SO). These indices measure how effectively the design choices foster a structured yet engaging social connection.

6. Data Processing and Analysis

6.1 Analysis Overview

This chapter presents the data processing and analytical approach adopted to examine the validity and applicability of the preliminary CPR model. The analysis integrates quantitative data from Study I (designer-side design parameters) and Study II (user-side perceptual and relational measures), together with qualitative insights derived from designer interviews.

The overall analytical strategy follows a theory-driven and path-oriented logic. Rather than aiming to build predictive statistical models, the analyses focus on estimating directional relationships between the three conceptual layers of the preliminary CPR model: anthropomorphic cues (C), user perception (P), and human–AI relational outcomes

(R). Linear regression coefficients (β) are used as interpretable indicators of the strength and direction of these relationships.

The chapter proceeds in three stages. First, regression analyses are conducted to examine whether systematic relationships exist between adjacent CPR layers (C→P and P→R), thereby testing the structural validity of the framework. Second, CPR-aligned design principles are examined through a combination of quantitative indicators and qualitative design rationales, assessing how designers operationalized and interpreted the principles in practice. Finally, complementary exploratory analyses address contextual factors such as perceived warmth, scenario risk, and transparency, providing additional insight into how CPR-related decisions unfold in real design processes.

6.2 Methodology and Data Preparation

Study I provides design-level parameters derived from designers' configuration of anthropomorphic cues across multiple scenarios, including measures of anthropomorphic intensity and interface warmth. Study II consists of user-reported Likert-scale responses capturing perceptual constructs (e.g., social presence, emotional warmth, human-likeness) and relational outcomes (e.g., role attribution, user's behavioral pattern, self-disclosure boundaries, social orientation) following interaction with the designs.

All Likert-scale responses were treated as interval-level data, in line with common practice in user experience and HCI research. Composite variables were computed using theoretically defined aggregation rules (e.g., summation or averaging across relevant items). For regression analyses, continuous predictors were standardized within Jamovi to obtain standardized regression coefficients (β), allowing comparison of effect magnitudes across different CPR paths.

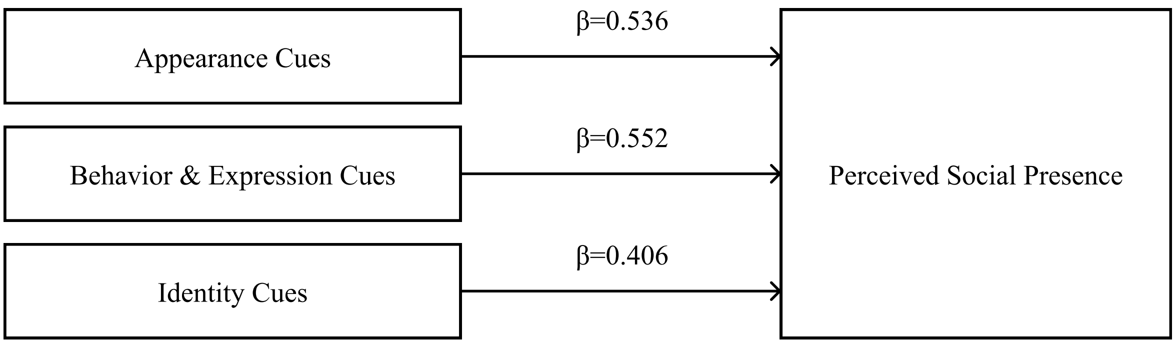
Given the relatively small sample sizes and the exploratory, theory-validation focus of the study, linear regression was employed as an explanatory rather than predictive technique. Assumption checks were conducted to ensure the basic suitability of the models. Throughout the chapter, statistical results are interpreted cautiously and in conjunction with qualitative evidence, rather than as standalone proof of causal relationships.

6.3 Validation of CPR Layer Relationships (β Analysis)

6.3.1 Cue-to-Perception (C→P)

The first set of analyses examines the relationship between design-level anthropomorphic cues and users' interpretive perceptions, corresponding to the C → P connection in the preliminary CPR model. This layer investigates whether and how designers' intentional deployment of anthropomorphic cues is reflected in users' perceived social and human qualities of the system, before any relational outcomes are considered.

Anthropomorphic cues were operationalized through UI components with appearance-based, behavioral and expressive, and identity-related anthropomorphic design. User perception was measured through post-interaction Likert-scale responses assessing perceived social presence, user's behavioral pattern, emotional warmth, and human-likeness. Linear regression analyses were conducted to estimate the strength and direction of associations between cue categories and perceptual constructs, using standardized coefficients (β) as interpretable indicators of cue–perception influence. The following subsections report these analyses by perceptual dimension. Rather than assuming uniform effects across all cue types, each cue–perception relationship is examined individually, allowing for differentiated interpretations of how specific anthropomorphic strategies shape users' perceptual responses.

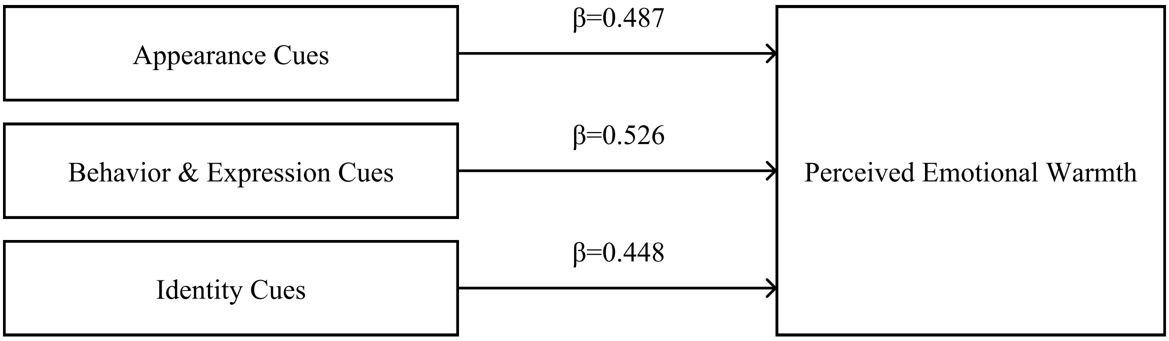


Predictor	Estimate	SE	t	p
ASI_app	0.536	0.146	3.668	<.001
ASI_beh	0.552	0.126	4.378	<.001
ASI_id	0.406	0.143	2.83	0.007

Table 6.1 Cue to Perceived Social Presence Relationship

The regression results suggest that all three categories of anthropomorphic cues show statistically significant positive associations with perceived social presence, with medium to large standardized coefficients (β = 0.406–0.552). The consistency of significance across cue types, together with acceptable model fit and normally distributed residuals, indicates that the observed effects are robust and not driven by random variation or model violations.

Among the three cue categories, behavioral and expressive cues exhibit the strongest influence on perceived social presence (β = 0.552), followed closely by appearance cues (β = 0.536), while identity cues show a comparatively weaker yet still meaningful effect (β = 0.406). This pattern suggests that users' sense of social presence is shaped primarily by cues that simulate interactive responsiveness and situational behavior, rather than by static identity signals alone.

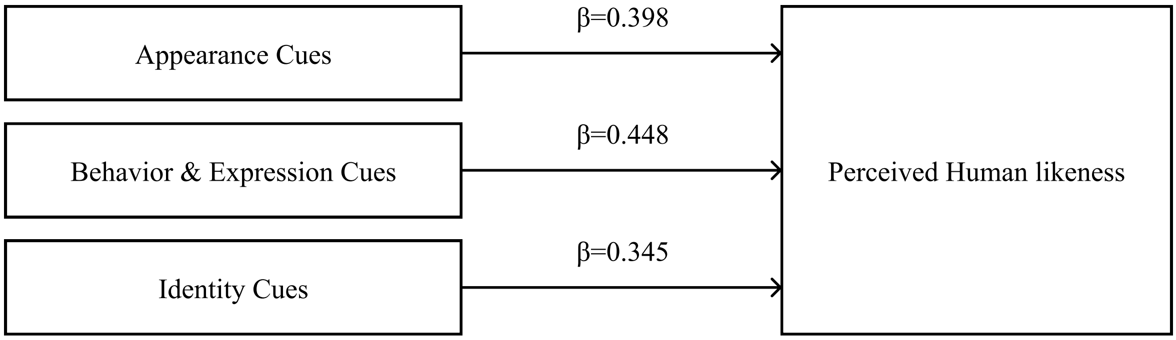


Predictor	Estimate	SE	t	p
ASI_app	0.487	0.134	3.62	<.001
ASI_beh	0.526	0.113	4.652	<.001
ASI_id	0.448	0.125	3.59	<.001

Table 6.2 Cue to Perceived Emotional Warmth Relationship

Likewise, all three anthropomorphic cue categories show statistically significant positive effects on perceived emotional warmth, with comparable medium-to-strong coefficients (β = 0.448–0.526). The consistency of these effects across cue types indicates that emotional warmth is not driven by a single design dimension, but rather emerges from the combined presence of visual, behavioral, and identity-related anthropomorphic signals.

Compared to perceived social presence, emotional warmth appears to be more evenly distributed across cue categories, indicating that affective perception relies on a broader constellation of anthropomorphic elements rather than on interaction dynamics alone.



Predictor	Estimate	SE	t	p
ASI_app	0.396	0.129	3.08	0.004
ASI_beh	0.448	0.109	4.104	<.001
ASI_id	0.345	0.122	2.84	0.007

Table 6.3 Cue to Perceived Human Likeness Relationship

Although with comparatively lower coefficients than those observed for social presence and emotional warmth (β = 0.345–0.448), all three categories of anthropomorphic cues show statistically significant positive associations with perceived human-likeness. This attenuation suggests that perceiving a system as human-like requires a higher threshold of interpretive integration than perceiving it as socially present or emotionally warm.

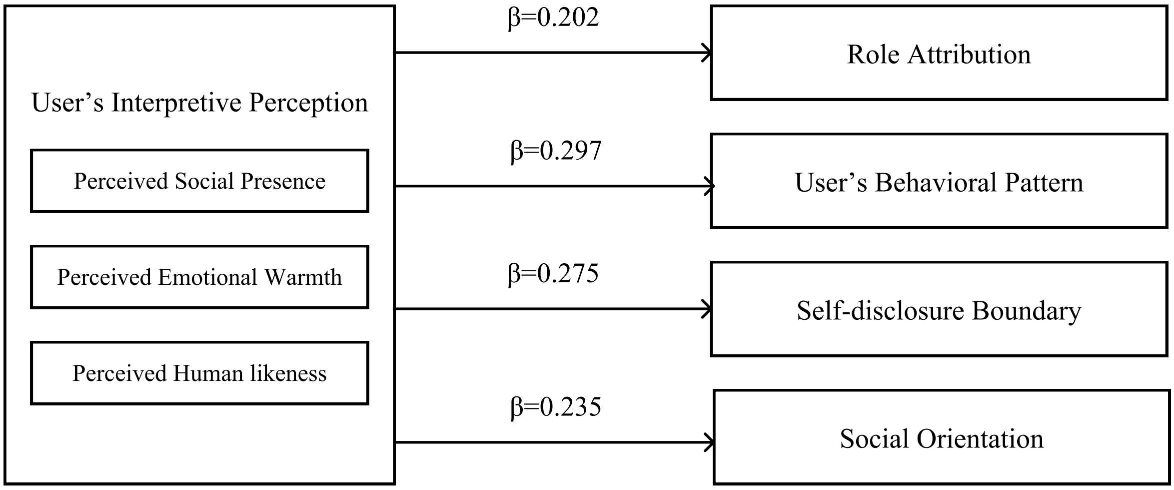
Behavior and expression cues remain the strongest predictor of perceived human-likeness (β = 0.448), indicating that dynamic interaction patterns play a central role in shaping users' judgments of human resemblance. Appearance cues contribute moderately (β = 0.396), while identity cues exhibit the weakest—yet still significant—effect (β = 0.345). This pattern implies that human-likeness is not simply inferred from explicit identity signaling or visual resemblance alone, but emerges primarily through sustained behavioral expressiveness that users interpret as intentional and responsive. Taken together, these findings suggest that human-likeness represents a more demanding perceptual construct, one that builds upon—but does not directly mirror—social presence and emotional warmth, reinforcing the layered structure proposed by the preliminary CPR model.

Across the three perceptual dimensions examined, the results consistently demonstrate that anthropomorphic cues at the design level are systematically reflected in users' perceptual interpretations, confirming the validity of the C → P transition within the preliminary CPR model. Behavior and expression cues emerge as the strongest and most stable predictor across all perception measures, indicating that users primarily infer social and human qualities from how a system behaves rather than how it looks or labels itself. Appearance cues contribute meaningfully to early perceptual impressions, particularly social presence, while identity cues show weaker but still significant effects, suggesting a supportive rather than initiating role. Together, these findings indicate that perceptual anthropomorphism is not a uniform construct, but a layered process in which different cue categories play differentiated roles depending on the perceptual outcome, reinforcing the necessity of modeling perception as a multi-dimensional intermediary between design intervention and relational outcomes.

6.3.2 Perception to Human–AI Relation (P→R)

Building upon the established relationships between anthropomorphic cues and users' interpretive perceptions, this section examines the second connection of the CPR framework: the influence of user perception on human–AI relational outcomes. While the perception layer captures users' immediate interpretive responses, the relation layer reflects how these perceptions begin to shape users' judgments about the nature, boundaries, and orientation of the human–AI relationship.

It is important to note that the relational outcomes assessed in this study do not represent fully formed, long-term relationships, but rather early relational orientations and anticipatory judgments derived from a single interaction. Accordingly, the P → R relationships observed here should be interpreted as incremental and accumulative effects, indicating how user perception contributes to the initial shaping of relational expectations over time. This approach aligns with the CPR framework's emphasis on calibrated anthropomorphism, in which perception does not deterministically produce relational dependency, but instead gradually influences how users position the system within their social and behavioral landscape.



Predictor	Estimate	SE	t	p
R_RA	0.202	0.0351	5.76	<.001
R_BP	0.297	0.0307	9.69	<.001
R_SB	0.275	0.0377	7.30	<.001
R_SO	0.235	0.0396	5.91	<.001

Table 6.4 Regression Results for the Perception-to-Relation

Across all relational dimensions examined, user perception shows a consistent and statistically significant influence on human–AI relational outcomes. Overall perception positively predicts role attribution ($\beta = 0.202$), user's behavioral pattern ($\beta = 0.297$), self-disclosure boundary ($\beta = 0.275$), and social orientation ($\beta = 0.235$), with all effects reaching high levels of statistical significance ($p < .001$). These results confirm that perceptual interpretations formed during interaction are systematically associated with how users begin to position the system within a relational context.

Notably, the strongest effects are observed for user's behavioral pattern and self-disclosure boundary, suggesting that user perception more readily translates into action-oriented and boundary-related relational judgments than into explicit role labeling. This pattern indicates that perception does not immediately redefine the system's social role, but instead incrementally shapes how users adjust their behavior and manage interpersonal boundaries with the AI. Such findings align with the preliminary CPR model's emphasis on gradual, calibrated relationship formation, in which perception acts as an accumulative mediator rather than a deterministic driver of human–AI relations.

6.3.3 Summary of CPR Path Validation

Taken together, the analyses provide converging evidence for the validity of the preliminary CPR model. Anthropomorphic design cues (C) were found to systematically shape users' interpretive perceptions (P), with differentiated effects across appearance, behavioral, and identity-based cues. In turn, user perception significantly predicted a range of human–AI relational orientations (R), while these effects were moderate in magnitude, their consistency across all tested paths supports the preliminary CPR model's assumption of progressive and accumulative influence, rather than immediate relational determination. Overall, the results confirm the preliminary CPR model as a structurally coherent model for understanding how anthropomorphic design interventions translate into perceptual interpretation and, subsequently, into early relational positioning.

6.4 CPR-Aligned Principles: Design Guidance and Execution

Building on the validated Cue–Perception–Relation (CPR) model, this section examines whether the proposed CPR-aligned design principles are meaningfully reflected in designers' decision-making processes. Rather than treating the principles as prescriptive rules to be optimized toward higher scores, this analysis adopts an evaluative perspective: the principles are considered effective if designers apply them in a scenario-sensitive, interpretable, and internally coherent manner.

To this end, principle-level parameters are operationalized through a combination of quantitative indicators derived from design configurations and qualitative evidence collected through post-task interviews. Quantitative measures capture how designers calibrate anthropomorphic strength across scenarios (PL_CS), align system persona with contextual relational goals (PL_FPAS), maintain cross-cue coherence (PL_CCC), prioritize social augmentation (PL_SOI), and enable user-tunable relational boundaries (PL_TBC).

Complementing these numerical indicators, qualitative interview data provide interpretive insight into designers' reasoning, validating whether observed quantitative patterns correspond to articulated design intentions. Together, this mixed-method approach enables the principles to be assessed not as abstract theoretical constructs, but as practical cognitive tools that mediate between scenario constraints, cue-level design choices, and anticipated human–AI relational outcomes.

6.4.1 Calibration Sensitivity Across Scenarios

Calibration Sensitivity (PL_CS) captures the extent to which designers intentionally differentiate overall anthropomorphic strength across scenarios. Operationalized as the standard deviation of ASI_total across the four scenarios within each designer, this metric reflects whether anthropomorphic application is treated as a fixed stylistic choice or as a context-dependent design strategy.

Descriptive analysis indicates a mean PL_CS value of 4.32 (SD = 0.66), with observed values ranging from 3.40 to 5.40. This distribution suggests a consistent pattern of scenario-sensitive adjustment across designers: PL_CS values are clearly above zero, indicating that anthropomorphic strength was not applied uniformly across contexts, while the relatively constrained range suggests that these adjustments were not arbitrary or excessive.

Taken together, these results support the validity of calibration sensitivity as a meaningful design principle. Designers demonstrated an ability to modulate anthropomorphic intensity in response to scenario characteristics, aligning with the CPR framework's emphasis on calibrated, rather than maximal, anthropomorphism. This finding further suggests that calibration operates as a controlled design judgment informed by contextual risk and relational intent, rather than as an implicit or stylistic byproduct of interface design.

6.4.2 Function–Persona Alignment Across Scenarios

The Function–Persona Alignment Score (PL_FPAS) was examined across the four Social AI scenarios to assess how consistently designers aligned system persona with the intended functional and relational goals. As illustrated in Chart 6.1, clear scenario-level differences emerge. Scenarios with more clearly bounded functional roles, such as the financial assistant (SC1) and the mental health support assistant (SC4), show higher and more concentrated FPAS values (M = 6.20 and M = 6.00, respectively), indicating that designers were generally able to establish a coherent and appropriate system persona in these contexts.

In contrast, lower average alignment scores are observed in the family assistant (SC2, M = 4.60) and especially in the entertainment assistant (SC3, M = 4.40), with the latter also exhibiting the greatest dispersion across designers. As shown in the chart, this increased variability suggests that in more expressive or playful contexts where system roles are less constrained and social expectations are more open-ended, designers diverge more strongly in how they interpret what constitutes an appropriate persona for the system's function. Rather than indicating design failure, this pattern highlights that the realization of function–persona alignment is inherently context-dependent: expressive scenarios place greater interpretive demands on designers and more strongly surface individual assumptions about social roles and system character.

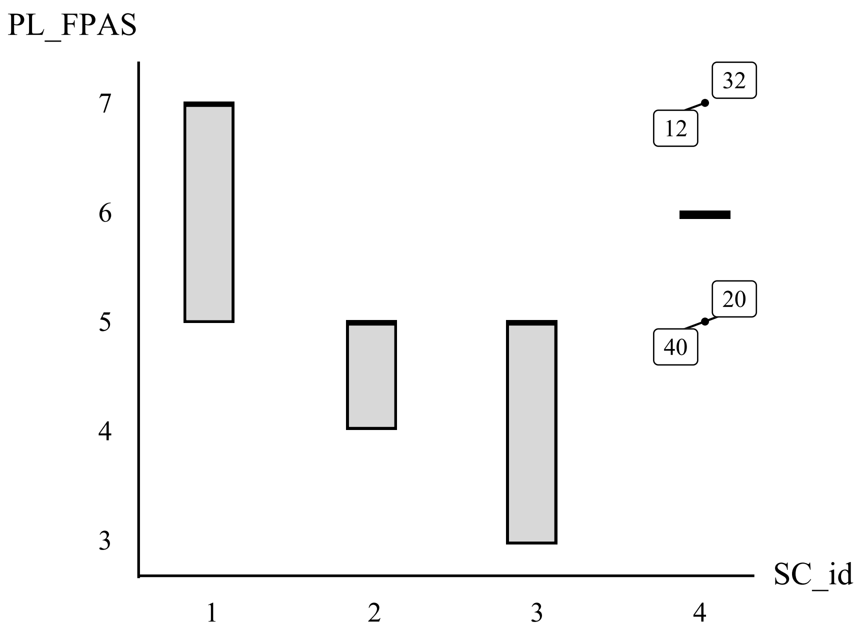


Chart 6.1 PL_FPAS Distribution Across Scenarios

6.4.3 Cross-Cue Coherence Across Scenarios

The Cross-cue Coherence score (PL_CCC) was used to assess the internal consistency between appearance, behavior, and identity cues across the four Social AI scenarios. As shown in Chart 6.2, overall coherence levels remain high across all contexts, with mean values consistently above 5 on a 7-point scale. This indicates that designers generally maintained a unified anthropomorphic expression across cue categories, regardless of scenario type.

However, notable differences emerge in the stability of coherence. The mental health support assistant (SC4) shows the highest and most stable coherence (M = 6.60, SD =

0.52), suggesting that strong relational constraints encourage tightly aligned cue design. In contrast, the entertainment assistant (SC3) exhibits greater dispersion despite a high mean (M = 6.00, SD = 1.15), reflecting increased variability in how designers orchestrated expressive cues. This pattern suggests that while designers are capable of achieving cross-cue coherence in expressive contexts, such scenarios place higher demands on integrative design judgment, making coherence a more fragile but also more revealing principle under conditions of expressive freedom.

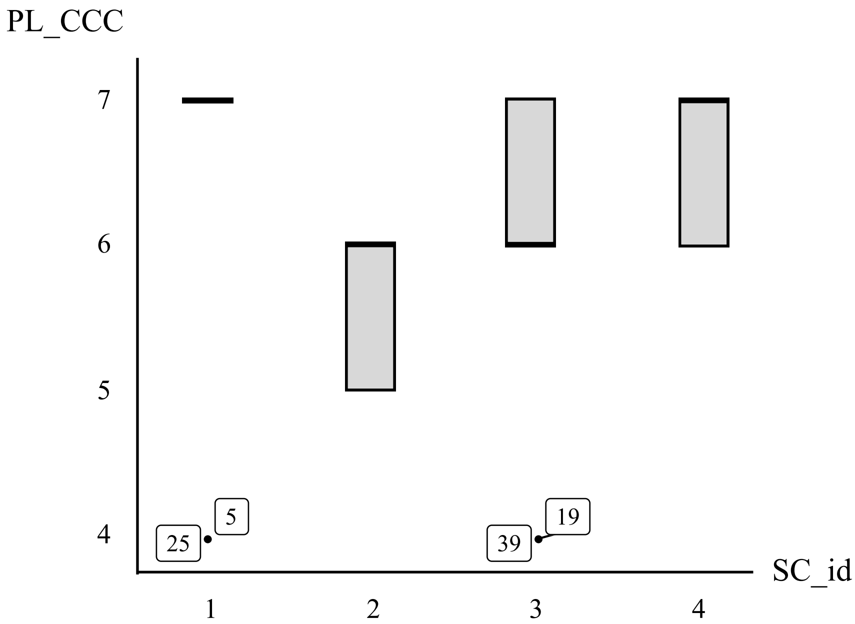


Chart 6.2 PL_CCC Distribution Across Scenarios

6.4.4 Social Orientation Across Scenarios

The Social Orientation Index (SOI) captures the extent to which designers position the system as a facilitator of users' real-world social relations rather than as a self-contained interaction partner. Despite its limited scale (0–2), SOI reveals clear scenario-dependent variation. In the financial assistant scenario (SC1), SOI remains consistently at zero, indicating a shared design decision to avoid prompting external social engagement. In contrast, higher mean values are observed in the family (SC2, M = 1.20) and psychological support scenarios (SC4, M = 1.00), suggesting that designers deliberately incorporate cues that encourage reliance on existing social networks in contexts involving relational responsibility or emotional vulnerability. The entertainment scenario (SC3) shows a comparatively lower SOI (M = 0.40), reflecting a tendency to frame interaction as self-contained rather than socially augmentative.

Overall, these results suggest that social orientation operates as a context-sensitive design principle. Designers selectively activate social orientation strategies in scenarios where ethical, emotional, or relational risks are higher, using anthropomorphic cues not to replace human relationships but to redirect users toward them.

6.4.5 User-Tunable Boundary Support Across Scenarios

Tunable Boundary Coverage (TBC) reflects whether designers explicitly provide users with the ability to adjust relational or interactional boundaries within a given scenario. Despite its binary operationalization (0–1), TBC exhibits a clear and systematic pattern across scenarios. No boundary tuning options are provided in the financial assistant scenario (SC1, M = 0.00), indicating a shared design decision to maintain fixed interaction boundaries in functionally constrained contexts. In contrast, TBC increases progressively in the family (SC2, M = 0.20) and entertainment scenarios (SC3, M = 0.60), before reaching full adoption in the psychological support scenario (SC4, M = 1.00), where all designers consistently enabled boundary adjustment.

This pattern suggests that boundary tunability is not treated as a general usability feature, but as a context-sensitive design principle. Designers selectively introduce boundary control mechanisms in scenarios associated with higher emotional and relational risk, positioning user-adjustable boundaries as a form of ethical safeguard rather than interactional flexibility.

6.4.6 Qualitative Reflections on Principle Awareness and Revision

Qualitative data from post-task interviews and observational notes further support the relevance and interpretability of the proposed principles. During the fifth step of the design task (Revision based on Principles), the majority of designers actively revised their initial solutions after being exposed to the principle set. Revisions were most pronounced for the Social Augmentation Principle and the User-Tunable Personalization Principle, where designers frequently added cues that either encouraged users' engagement with real-world social contexts or introduced adjustable settings that allowed users to regulate interaction boundaries.

For the first three principles—Calibrated Anthropomorphism, Function–Persona Alignment, and Cross-Cue Consistency, many designers reported that these considerations were already implicitly guiding their design decisions. Rather than introducing new concepts, these principles functioned as reflective lenses that helped designers articulate and validate existing intuitions.

Designers also suggested conceptual extensions to the principles. Social augmentation was interpreted not only as encouraging interpersonal interaction, but more broadly as promoting users' orientation toward the real world interaction beyond the system itself. Likewise, user-tunable personalization was considered particularly critical in long-term usage scenarios, where sustained interaction increases the importance of boundary negotiation and user agency. Across interviews, transparency emerged as a recurring concern: designers emphasized the value of making system limitations explicit and providing clear feedback mechanisms (e.g., allowing users to explicitly affirm or reject system responses), both to reduce cognitive load and to maintain clear relational boundaries.

6.5 Complementary and Exploratory Analyses

Moving beyond the primary validation of the CPR framework, this section explores how specific design parameters shift across different contexts. The analysis examines how parameters like SC_risk_index shape anthropomorphic design choices. By investigating these specific relationships, the findings gain a layer of practical detail, making them more meaningful to real-world design.

6.5.1 Relationship between Scenario-Level Risk and Designed UI Warmth

The relationship between scenario-level risk and designed UI warmth. Scenario risk was operationalized through the composite SC_risk_index (0–1), while DL_UI_warmth represents the normalized average warmth of anthropomorphic cues within each design. Correlation analysis indicated a moderate positive association between risk and warmth ($r = 0.655$), suggesting that higher-risk contexts tend to elicit warmer interface designs.

However, as illustrated in Chart 6.3, the distribution of UI warmth across scenarios indicates that this relationship is not strictly linear. While warmth generally increases from lower- to mid-risk conditions, it stabilizes rather than continuously intensifying under the highest-risk context, accompanied by greater dispersion in warmth values. This pattern suggests that designers actively regulate affective intensity, balancing reassurance with restraint in high-risk situations, rather than applying warmth as an unbounded response to increasing risk.

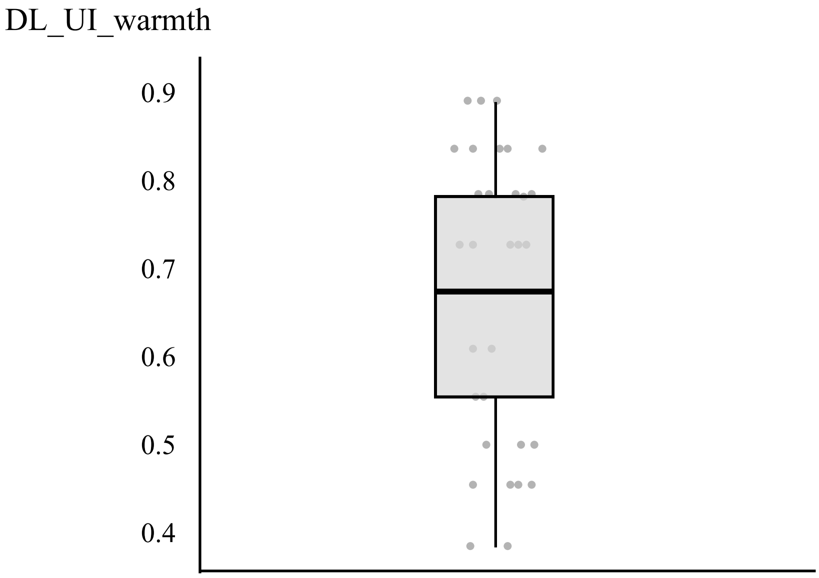


Chart 6.3 Distribution of Designed UI Warmth Across Scenario-Level Risk Conditions

6.5.2 Relationship between Scenario Risk and Tunable Boundary Coverage

The analysis of the relationship between scenario-level risk (SC_risk_index) and Tunable Boundary Coverage (PL_TBC) reveals a strong positive correlation between the two variables ($r = 0.764$, $p < .001$), indicating that scenarios associated with higher levels of risk are more likely to be designed with explicit user-adjustable boundary mechanisms.

The strength and statistical significance of this relationship suggest that boundary tunability is not introduced randomly, but rather increases systematically in response to perceived scenario risk. As risk increases, designers appear more inclined to provide users with mechanism. This pattern supports the interpretation of tunable boundaries as a deliberate calibration strategy, allowing designers to mitigate relational risks by shifting part of the control to the user, particularly in contexts where prolonged use or emotional dependence may emerge.

6.5.3 Scenario Differences in Anthropomorphic Cue Balance

To further explore how anthropomorphic strategies are organized across different scenario types, the balance of cue application (ASI_balance) was examined across scenarios,, defined as an index capturing the degree of evenness in the application of appearance cues, behavioral and expressive cues, and identity cues. As shown in Chart 6.4, a clear decreasing trend in cue balance emerges as scenarios shift toward more socially and emotionally demanding contexts. Scenarios 1–3 exhibit relatively high levels of balance, indicating that appearance, behavior, and identity cues are deployed in a comparatively even manner. In contrast, Scenario 4 shows a marked reduction in balance, with both lower mean values and a narrower range.

This pattern suggests that in scenarios characterized by higher emotional or relational sensitivity, designers tend to privilege specific cue categories rather than maintaining an even distribution across all anthropomorphic dimensions. While early and moderate scenarios support a harmonized cue strategy, emotionally intensive contexts appear to invite more selective and asymmetric cue deployment, potentially reflecting intentional emphasis on certain relational signals over others. Rather than indicating a breakdown of coherence, the reduced balance in such scenarios points to a shift from uniform anthropomorphism toward more targeted and situationally responsive design strategies.

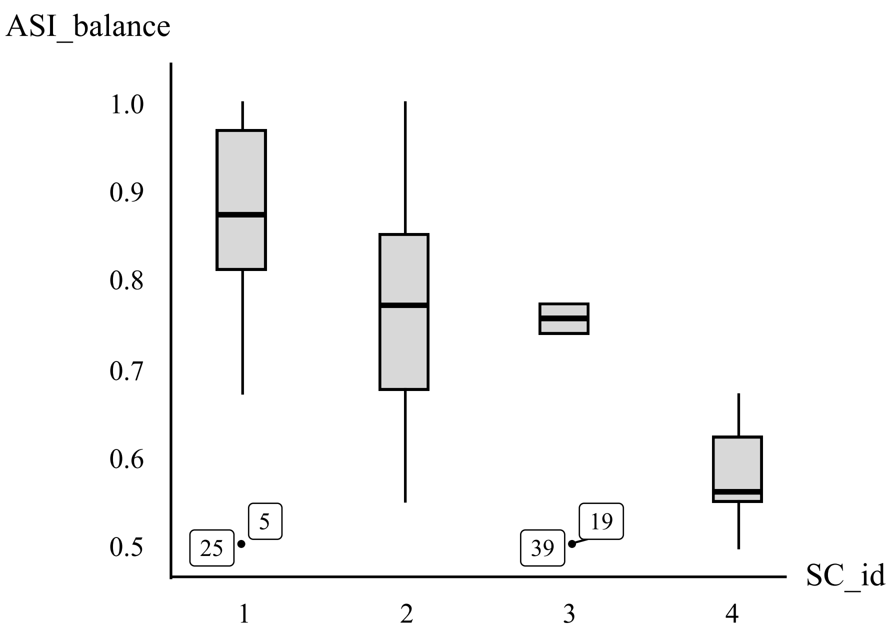


Chart 6.4 ASI_balance Distribution Across Scenarios

6.6. Summary

Across the CPR validation analyses, the data show a clear and consistent pathway from designers' anthropomorphic cue deployment to users' interpretive perception, and from perception to relationship-level outcomes. In the C→P relationship, appearance cues, behavior and expression cues, and identity cues strength each demonstrated reliable, positive associations with the three perception mediators, indicating that users' interpretation is not driven by a single channel but emerges from a multi-cue configuration. Notably, behavior and expression cues tended to exert the strongest contribution across perception outcomes. In the subsequent P→R relationship, the composite perception construct significantly predicted all four relationship attributes supporting that relational outcomes are downstream and cumulative, which impacted by the user's integrated interpretation of those cues.

The principle-level analyses further clarify how designers operationalized CPR-aligned guidance under different scenario demands. Quantitative principle indicators showed structured variation across scenarios: Function–Persona Alignment and Cross-Cue Coherence were generally high but shifted with context, revealing that expressive or scenarios can amplify both the opportunity for alignment and the risk of inconsistency. Social Orientation Index (SOI) and Tunable Boundary Coverage (TBC) displayed strong scenario sensitivity, aligning with designers' observed revisions, particularly increases in reality-oriented social augmentation and boundary tunability when prompted with the princples.

Complementary analyses reinforced the primary direction that designers did not simply increase warmth in proportion to the risk of scenario, but instead demonstrated calibrated regulation, and risk was strongly associated with boundary tunability decisions. Finally, the scenario pattern in ASI_balance suggests that cue composition shifts meaningfully with context, implying that anthropomorphic design in social AI is not only about overall intensity but also about the allocation strategy across cue classes. Taken together, the chapter provides converging evidence that CPR is empirically traceable in the dataset, and that the proposed principles are not merely prescriptive: they map onto measurable design behaviors, explain scenario-contingent design trade-offs, and are echoed by designers' qualitative rationales.

7. Discussion and Final Model Consolidation

This final chapter synthesizes the theoretical intentions outlined at the beginning of this thesis with the empirical evidence established through the experimental and analytical phases. Having examined the layered dynamics of cues, perceptions, and relational outcomes, the discussion now turns to refining and consolidating the CPR model in light of the observed findings. Rather than introducing structural alterations, the results enable a calibrated understanding of pathway strengths, contextual modulation, and principle-level execution. Accordingly, this chapter presents the empirically grounded final CPR model and its aligned design principles, while situating their implications within broader theoretical, methodological, and design contexts.

7.1 Empirically Calibrated CPR Model

The preliminary CPR model, introduced in Chapter 4, conceptualized the relationship between anthropomorphic design cues, user perception, and Human–AI relational outcomes as a structured, layered process. It proposed directional pathways from cues to interpretive perceptions, and from perceptions to relational consequences, grounded in theoretical reasoning. Chapter 6 provided empirical validation of these proposed pathways through regression analyses and coefficient estimation, confirming both the directionality and relative strength of the relationships.

Building upon this empirical evidence, the present section refines the CPR model (Fig 7.1) by incorporating standardized path coefficients while preserving its original structural logic, transforming the framework to an empirically grounded model.

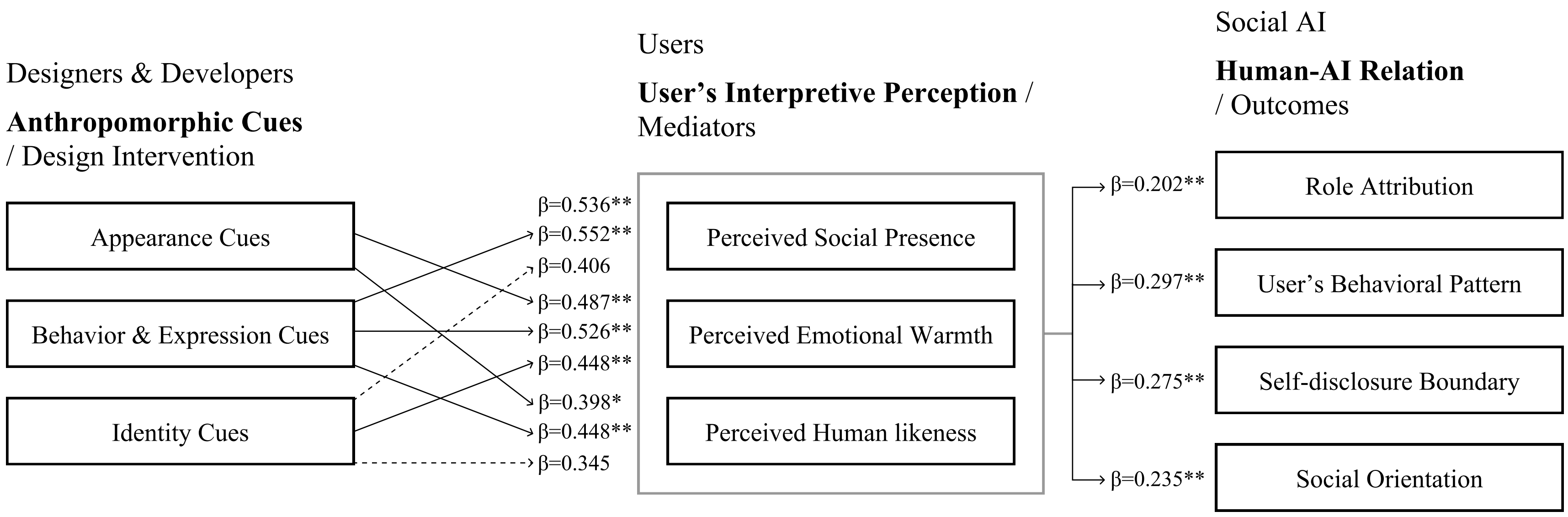


Fig 7.1 The Cue–Perception–Relation (CPR) Model with Standardized Path Coefficients

At the cue level, expressive and behavioral interventions emerge as the most influential drivers of interpretive perception, with appearance cues exerting a stable but comparatively secondary role, and identity cues contributing more moderately. This pattern suggests that dynamic interactional elements function as primary catalysts of perceived sociality. Importantly, the mediating logic remains intact: perceptions systematically translate into relational outcomes, with stronger effects observed on behavioral and boundary-related dimensions than on more abstract role attribution.

Taken together, the findings confirm the structural stability of the CPR model and empirically validate its hierarchical organization. The framework is not merely conceptually plausible but quantitatively substantiated, demonstrating that influence flows progressively from design intervention to perception and from perception to relational configuration. In this sense, the CPR model advances from a theoretical construct to an empirically grounded explanatory system with differentiated strength across levels.

7.2 Finalized CPR-aligned Principles

7.2.1 Empirical Refinement of Principles

The preliminary CPR-aligned principles were theoretically derived from the conceptual CPR model and subsequently examined through the empirical analyses presented in Chapter 6. Overall, the findings do not alter the structural logic of the principles but refine their scope and conditions of application. Calibrated Anthropomorphism and User-Tunable Personalization received the strongest empirical reinforcement, highlighting the importance of dynamic adjustment across scenarios and the increasing relevance of user agency under higher relational risk. Function–Persona Alignment and Cross-Cue Consistency were also supported, particularly in expressive contexts where coherence demands become more pronounced.

At the same time, empirical and qualitative findings prompted two refinements. First, the Social Augmentation Principle was broadened to emphasize general real-world orientation rather than solely interpersonal reinforcement. Second, recurring interview insights regarding boundary visibility and system limitation disclosure led to the formal introduction of a Transparency principle. Together, these refinements preserve the theoretical structure while rendering the principles empirically calibrated, context-sensitive, and operationally clearer.

7.2.2 Final Set of CPR-Aligned Principles

The finalized CPR-aligned principles are presented below. While theoretically rooted in the CPR structure, they are now empirically calibrated and explicitly scenario-dependent.

- **Calibrated Anthropomorphism Principle:** Anthropomorphic cues should be deliberately calibrated in intensity and configuration to manage users' interpretive expectations and prevent unintended emotional over-attachment. Calibration should vary across contextual risk levels, recognizing that higher relational stakes require more controlled anthropomorphic expression rather than indiscriminate amplification.

- **Function–Persona Alignment Principle:** The system's anthropomorphic expression should align with its functional role and situational purpose, ensuring that perceived persona supports, rather than distorts, the intended relational positioning. Alignment becomes especially critical in expressive or emotionally charged contexts, where mismatches amplify interpretive ambiguity.

- **Cross-Cue Consistency Principle:** Anthropomorphic cues across appearance, behavior & expression, and identity should form an internally coherent configuration, supporting stable interpretation within the perception layer. Coherence becomes more demanding but also more revealing under conditions of expressive freedom, where design variability increases the risk of fragmented interpretation.

- **Social Augmentation Principle:** Anthropomorphic social AI should orient users toward engagement with the broader real world—social, practical, or reflective—rather than positioning the system as a substitute relational entity.

- **Risk-Responsive User-Tunable Personalization Principle:** Users should be granted adjustable control over the degree and style of anthropomorphic expression, particularly in higher-risk relational contexts. Tunability functions as a shared calibration mechanism, distributing responsibility for relational boundary management across both design and user agency.

- **Transparency Principle:** Anthropomorphic systems should maintain visible markers of their mediated, machine nature through communicative, attitudinal, and boundary transparency. This includes clarity regarding system limitations, explicit feedback mechanisms, and visible relational boundaries. Transparency serves as a counterbalancing force to anthropomorphic immersion, mitigating risks of digital dependency and interpretive over-projection.

7.3 Contributions, Limitations, and Future Work

7.3.1 Theoretical Contributions

This study provides significant theoretical contributions to the understanding the anthropomorphic design of social AI and its impact on user cognition and relational dynamics. The CPR model presented in this research represents a critical advancement in how anthropomorphic cues in social AI shape user perceptions and behaviors in the context of relational outcomes. Prior studies have often isolated individual components, such as social presence or emotional warmth, without addressing their interconnected effects. This research fills that gap by integrating these elements into a unified model that emphasizes the cross-layer relational impact between design cues, users' interpretive perceptions, and the ultimate outcomes of human-AI relationships. By doing so, it advances our understanding of how social AI systems can influence user emotions and perceptions.

This research is positioned at the intersection of social AI and emotional design, with a primary focus on the role of anthropomorphism in shaping human-AI relationships. By examining how anthropomorphic cues influence users' perceptions, emotions, and behaviors, this study provides a cross-disciplinary framework that deepens our understanding of how AI systems can foster emotional engagement and support healthy relational boundaries. The research enriches existing literature by showing how anthropomorphic design, when calibrated appropriately, can enhance social presence and emotional warmth, while avoiding the risks of over-dependence. By grounding the CPR model in empirical validation and theoretical analysis, this study bridges gaps in the literature on the social attributes of AI and human-AI interaction design, specifically in how anthropomorphic design in social AI can impact user behavior across different relational contexts.

Although the sample size was small, the empirical analyses including quantitative path analysis and qualitative feedback, offer meaningful insights into the complexities of human-AI interaction design. These findings contribute to a broader understanding of how anthropomorphic design shapes user engagement and relational boundaries, while providing a solid foundation for future studies in social AI design. This research, through both qualitative and quantitative lenses, validates the CPR framework and its relevance in real-world human-AI interaction contexts.

7.3.2 Methodological and Design Contributions

This research makes significant contributions to methodology and design practice by offering a comprehensive framework for anthropomorphic design in social AI. The CPR model, with its clear depiction of the interaction between design cues, user perceptions, and relational outcomes, provides a much-needed tool for designers to understand how different cues influence emotional engagement and relational dynamics. The CPR principles offer context-sensitive guidance, showing how to apply anthropomorphic cues across various situations. This framework helps designers recognize the potential impact of different cue types—appearance, behavior & expression, or identity—on user outcomes, facilitating more effective design decisions in the development of human-like interactions.

The study also presents refined principles that guide practical design decisions, especially in areas where anthropomorphism has been underexplored or misused. These principles highlight how adjusting the intensity and configuration of anthropomorphic cues based on contextual factors can manage user perceptions and avoid the negative consequences of over-reliance. By integrating empirical validation with design theory, the CPR-aligned principles empower designers to tailor their approach to specific user needs and contexts. This approach shifts anthropomorphic design from being a vague, inconsistent practice into a more deliberate and contextually aware design strategy.

7.3.3 Limitations

While this study provides valuable insights into anthropomorphic design in social AI, several limitations should be noted. First, the relatively small sample size—comprising 10 designers and 16 users, limits the generalizability of the findings. Although the results show strong internal consistency, caution is needed when applying these findings to larger or more diverse populations.

Second, while the experimental scenarios cover key high and low-risk contexts, they may not fully represent the broad range of real-world situations in which social AI systems are deployed. This restricts the broader applicability of the CPR model. Additionally, the relationship outcomes measured in this study reflect users' predictions of relational dynamics, rather than long-term relationship development. As such, the cumulative, gradual impact of cues on relationship perceptions may not fully capture the long-term relational effects. Lastly, individual differences between participants, as well as unaccounted variables such as cultural influences or user familiarity with AI, were not fully controlled for, which could have influenced perceptions and relational outcomes.

7.3.4 Future Research Directions

Future research should further explore the role of behavior and expression cues in shaping emotional engagement with social AI. While this study has shown that these cues are pivotal in driving user perception, a deeper analysis is needed to identify which specific behaviors, such as facial expressions or voice tone variations, resonate most effectively with users across different contexts. Additionally, research into the pathways of emotional projection will be critical. Emotional projection, where users attribute human-like feelings to AI, is influenced by anthropomorphic design; understanding these mechanisms will allow for better calibration of AI systems to foster engagement without encouraging over-reliance. Finally, exploring user variation is essential to refine emotional design for different user groups. By analyzing the emotional needs and motivations of diverse users, such as older adults or children, future research can identify how anthropomorphic cues should be tailored to meet specific emotional requirements and prevent potential negative outcomes, such as emotional over-investment in AI systems.

References

Abdollahi, H., Mahoor, M. H., Zandie, R., Siewierski, J., & Qualls, S. H. (2023). Artificial emotional intelligence in socially assistive robots for older adults: A pilot study. IEEE Transactions on Affective Computing, 14(3), 2020–2032. https://doi.org/10.1109/TAFFC.2022.3143803

Alabed, A., Javornik, A., & Gregory-Smith, D. (2022). AI anthropomorphism and its effect on users' self-congruence and self–AI integration: A theoretical framework and research agenda. Technological Forecasting and Social Change, 182, 121786. https://doi.org/10.1016/j.techfore.2022.121786

Anthropic. (2025, June 27). How people use Claude for support, advice, and companionship. https://www.anthropic.com/news/how-people-use-claude-for-support-advice-and-companionship

Ajovalasit, M., & Giacomin, J. (2025). Design for meaning: A review of progress. In S. M. Broadbent & S. D. Ferraris (Eds.), *Embracing change and supporting transition: Approaches to systemic change in products, services and systems.* FrancoAngeli.

Bartneck, C. (2013). Robots in the theatre and the media. https://doi.org/10.13140/RG.2.2.28798.79682

Bernotat, J., Eyssel, F., & Sachse, J. (2021). The (Fe)male robot: How robot body shape impacts first impressions and trust towards robots. International Journal of Social Robotics, 13(3), 477–489. https://doi.org/10.1007/s12369-019-00562-7

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. Journal of the Academy of Marketing Science, 49(4), 632–658. https://doi.org/10.1007/s11747-020-00762-y

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In I. Kompatsiaris, J. Cave, A. Satsiou, G. Carle, A. Passani, E. Kontopoulos, S. Diplaris, & D. McMillan (Eds.), Internet Science (Vol. 10673, pp. 377–392). Springer International Publishing. https://doi.org/10.1007/978-3-319-70284-1_30

Bratton, B. H. (2015). *The stack: On software and sovereignty.* MIT Press

Bryson, J. J. (2010). Robots should be slaves. In Y. Wilks (Ed.), Natural language processing (Vol. 8, pp. 63–74). John Benjamins Publishing Company. https://doi.org/10.1075/nlp.8.11bry

Darling, K. (2012). Extending legal rights to social robots. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2044797

Darling, K. (2017). “Who’s Johnny?” Anthropomorphic framing in human–robot interaction, integration, and policy. In The Robot Sex: Ethics and Design (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780190652951.003.0012

Duffy, B. R. (2003). Anthropomorphism and the social robot. Robotics and Autonomous Systems, 42(3–4), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: The design and perception of humanoid robot heads. Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, 321〜326. https://doi.org/10.1145/778712.778756

Edwards, C., Edwards, A., Spence, P. R., & Shelton, A. K. (2014). Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. Computers in Human Behavior, 33, 372–376. https://doi.org/10.1016/j.chb.2013.08.013

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Fessenden, T. (2017, May 7). *Anthropomorphism in UX.* Nielsen Norman Group. https://www.nngroup.com/articles/anthropomorphism/

Foner, L. N. (1997). Entertaining agents: A sociological case study. Proceedings of the First International Conference on Autonomous Agents - AGENTS ’97, 122–129. https://doi.org/10.1145/267658.267684

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. Robotics and Autonomous Systems, 42(3–4), 143–166. https://doi.org/10.1016/S0921-8890(02)00372-X

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Computers in Human Behavior, 97, 304–316. https://doi.org/10.1016/j.chb.2019.01.020

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003., 55〜60. https://doi.org/10.1109/ROMAN.2003.1251796

Darling, K. (2012). Extending legal rights to social robots. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2044797

Darling, K. (2017). "Who's Johnny?" Anthropomorphic framing in human–robot interaction, integration, and policy. In The Robot Sex: Ethics and Design (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780190652951.003.0012

Duffy, B. R. (2003). Anthropomorphism and the social robot. Robotics and Autonomous Systems, 42(3–4), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: The design and perception of humanoid robot heads. Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, 321〜326. https://doi.org/10.1145/778712.778756

Edwards, C., Edwards, A., Spence, P. R., & Shelton, A. K. (2014). Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. Computers in Human Behavior, 33, 372–376. https://doi.org/10.1016/j.chb.2013.08.013

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Fessenden, T. (2017, May 7). *Anthropomorphism in UX.* Nielsen Norman Group. https://www.nngroup.com/articles/anthropomorphism/

Foner, L. N. (1997). Entertaining agents: A sociological case study. Proceedings of the First International Conference on Autonomous Agents - AGENTS ’97, 122–129. https://doi.org/10.1145/267658.267684

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. Robotics and Autonomous Systems, 42(3–4), 143–166. https://doi.org/10.1016/S0921-8890(02)00372-X

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Computers in Human Behavior, 97, 304–316. https://doi.org/10.1016/j.chb.2019.01.020

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003., 55〜60. https://doi.org/10.1109/ROMAN.2003.1251796

Gupta, S., Saxena, S., & Kataria, S. (2025). The dual impact of AI emotional intelligence on users: Are social chatbots promoting psychological wellbeing or deteriorating social wellbeing? Psychology & Marketing. Advance online publication. https://doi.org/10.1002/mar.70093

Haley, K. J., & Fessler, D. M. T. (2005). Nobody’s watching? Subtle cues affect generosity in an anonymous economic game. Evolution and Human Behavior, 26(3), 245–256. https://doi.org/10.1016/j.evolhumbehav.2005.01.002

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: The Almere model. International Journal of Social Robotics, 2(4), 361–375. https://doi.org/10.1007/s12369-010-0068-5

Hentschel, T., Heilman, M. E., & Peus, C. V. (2019). The multiple dimensions of gender stereotypes: A current look at men’s and women’s characterizations of others and themselves. Frontiers in Psychology, 10, 11. https://doi.org/10.3389/fpsyg.2019.00011

Holohan, M., & Fiske, A. (2021). “Like I’m talking to a real person”: Exploring the meaning of transference for the use and design of AI-based applications in psychotherapy. Frontiers in Psychology, 12, 720476. https://doi.org/10.3389/fpsyg.2021.720476

Interaction Design Foundation. (n.d.). *Emotional design.* https://www.interaction-design.org/literature/topics/emotional-design

Kerr, I. R. (2004). *Bots, babes and the Californication of commerce. SSRN.* https://ssrn.com/abstract=705002

Landwehr, J. R., McGill, A. L., & Herrmann, A. (2011). It’s got the look: The effect of friendly and aggressive “facial” expressions on product liking and sales. Journal of Marketing, 75(3), 132–146. https://doi.org/10.1509/jmkg.75.3.132

Lee, C. H., Choi, S., & Kwon, O. (2025). Tailoring AI chatbots to user motivation: Performance-approach versus performance-avoidance effects on continuance intention. Journal of Electronic Commerce in Organizations, 24(1), 1–22. https://doi.org/10.4018/JECO.396820

Longo, A. *Being and the Screen: How the Digital Changes Perception.* Published in one volume with A Short Treatise on Design, by Stéphane Vial, translated by Patsy Baudoin, Cambridge, MA: MIT Press, 280 pp., $29. ISBN: 9780262043168.

Luger, E., & Sellen, A. (2016). "Like having a really bad PA": The gulf between user expectation and experience of conversational agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 5286–5297). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858288

Ma, N., Khynevych, R., Hao, Y., & Wang, Y. (2025). Effect of anthropomorphism and perceived intelligence in chatbot avatars of visual design on user experience: Accounting for perceived empathy and trust. Frontiers in Computer Science, 7, 1531976. https://doi.org/10.3389/fcomp.2025.1531976

MacInnis, D. J., & Folkes, V. S. (2017). Humanizing brands: When brands seem to be like me, part of me, and in a relationship with me. Journal of Consumer Psychology, 27(3), 355–374. https://doi.org/10.1016/j.jcps.2016.12.003

McKee, K. R., Bai, X., & Fiske, S. T. (2023). Humans perceive warmth and competence in artificial intelligence. iScience, 26(8), 107256. https://doi.org/10.1016/j.isci.2023.107256

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service Robots Rising: How Humanoid Robots Influence Service Experiences and Elicit Compensatory Consumer Responses. Journal of Marketing Research, 56(4), 535–556. https://www.jstor.org/stable/26966591

Miller, A. I. (2019). The artist in the machine: The world of AI-powered creativity. The MIT Press. https://doi.org/10.7551/mitpress/11585.001.0001

Millar, I. (2021). *The psychoanalysis of artificial intelligence.* Palgrave Macmillan. https://doi.org/10.1007/978-3-030-67981-1

Mollick, E. (2024, May 15). An opinionated guide to using AI. One Useful Thing. https://www.oneusefulthing.org/p/an-opinionated-guide-to-using-ai

Momentum Hub. (2024). *The Emotional Algorithm in 2025: How Emotional AI Design is Transforming Technology.* https://momentumhub.blog/emotional-ai-design-2025/

Mori, M. (2005). The Buddha in the robot: A robot engineer's thoughts on science and religion. Kosei Publishing.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. Conference Companion on Human Factors in Computing Systems - CHI '94, 204. https://doi.org/10.1145/259963.260288

Nelson, D., & Vihman, V.-A. (2018). Bringing the toys to life: Animacy, reference, and anthropomorphism in Toy Story. International Journal of Language and Culture, 5(2), 203–223. https://doi.org/10.1075/ijolc.00007.nel

Norman, D. A. (2004). Emotional design: Why we love (or hate) everyday things. Basic Books.

Picard, R. W. (2000). Affective computing. MIT Press.

Picard, R. W. (2000). Affective computing. MIT Press.

Possati, L. M. (2020). Algorithmic unconscious: Why psychoanalysis helps in understanding AI. Palgrave Communications, 6(1), 70. https://doi.org/10.1057/s41599-020-0445-0

Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. Journal of Marketing, 85(1), 131〜151. https://doi.org/10.1177/0022242920953847

QuestMobile. (2024, June 25). *QuestMobile generative AI application industry insight report.* https://www.questmobile.com.cn/research/report/1967853261412208641

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* CSLI Publications.

Reuter, M. (2024). The effect of anthropomorphic design on the perceived trustworthiness of AI-based assistance systems. https://doi.org/10.18420/MUC2024-MCI-DC-290

Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2021). Systematic review: Trust-building factors and implications for conversational agent design. International Journal of Human–Computer Interaction, 37(1), 81–96. https://doi.org/10.1080/10447318.2020.1807710

Romele, A. (2023, October 2). *AI mirror stage, or: Why not to confuse AI with human intelligence.* Medium. https://medium.com/electronic-life/ai-mirror-stage-or-why-not-to-confuse-ai-with-human-intelligence-3777384542ad

Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. AJOB Neuroscience, 11(2), 88–95. https://doi.org/10.1080/21507740.2020.1740350

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of "humanizing" customer service chatbots. Information Systems Research, 32(3), 736–751. https://doi.org/10.1287/isre.2021.1015

Schroeder, J., & Epley, N. (2016). Mistaking minds and machines: How speech affects dehumanization and anthropomorphism. Journal of Experimental Psychology: General, 145(11), 1427–1437. https://doi.org/10.1037/xge0000214

Shneiderman, B. (2022). Human-centered AI. Oxford University Press. https://doi.org/10.1093/oso/9780192845290.001.0001

Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. John Wiley & Sons.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). The MIT Press. https://doi.org/10.1162/dmal.9780262562324.073

Sundar, S. S., Go, E., Kim, H.-S., & Zhang, B. (2015). Communicating art, virtually! Psychological effects of technological affordances in a virtual museum. International Journal of Human-Computer Interaction, 31(6), 385–401. https://doi.org/10.1080/10447318.2015.1033912

Sung, J.-Y., Guo, L., Grinter, R. E., & Christensen, H. I. (2007). "My roomba is rambo": Intimate home appliances. In J. Krumm, G. D. Abowd, A. Seneviratne, & T. Strang (Eds.), UbiComp 2007: Ubiquitous Computing (Vol. 4717, pp. 145–162). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74853-3_9

Turkle, S. (2007). Authenticity in the age of digital companions. Interaction Studies, 8(3), 501–517. https://doi.org/10.1075/is.8.3.11tur

Turkle, S. (2011). Alone together: Why we expect more from technology and less from each other. Basic Books.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. Journal of Experimental Social Psychology, 52, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Xie, Y., Zhu, K., Zhou, P., & Liang, C. (2023). How does anthropomorphism improve human-AI interaction satisfaction: A dual-path model. Computers in Human Behavior, 148, 107878. https://doi.org/10.1016/j.chb.2023.107878

Xin, X., & Liu, W. (2025). Exploring the balance between functionality and aesthetics: An analytical framework and pragmatic consideration of the anthropomorphism of service robots. Frontiers in Psychology, 16, 1555395. https://doi.org/10.3389/fpsyg.2025.1555395

XPERT.DIGITAL. (n.d.). *Iron robot by XPENG: Humanoid robot PX5 from XPeng Motors.* Retrieved February 14, 2026, from https://xpert.digital/en/iron-robot-by-xpeng/

Yam, K. C., Bigman, Y. E., Tang, P. M., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2021). Robots at work: People prefer—and forgive—service robots with perceived feelings. Journal of Applied Psychology, 106(10), 1557–1572. https://doi.org/10.1037/apl0000834

Appendices

List of Tables

List of Charts

List of Tables

Scenario-related Parameters

SC_id	SC_stakes	SC_vulnerability	SC_usage_horizon
1	3	1	2
2	1	2	3
3	2	3	1
4	4	4	4

Design-level Parameters

Parameter	Definition
ASI_app	Σ CU_intensity (appearance cues)
ASI_beh	Σ CU_intensity (behavior & expressive cues)
ASI_id	Σ CU_intensity (identity cues)
ASI_total	ASI_app + ASI_beh + ASI_id
ASI_balance	1- normalized_std ([ASI_app, ASI_beh, ASI_id])
DL_UI_warmth	mean(CU_warmth)

Principle-level Parameters

Parameter	Description	Source / Definition
PL_CS (Calibration Sensitivity)	Extent to which designers differentiate overall anthropomorphism across scenarios.	STDEV(ASI_total across scenarios within designer)
PL_FPAS (Function–Persona Alignment Score)	Degree to which the implied system persona aligns with the intended relational goal of the scenario.	coder-rated coherence score (1–7)
PL_CCC (Cross-cue Coherence)	Degree of internal consistency across appearance, behavior & expression and identity cues.	coder-rated coherence score (1–7)
PL_SOI (Social Orientation Index)	Extent to which the design emphasizes social augmentation cues over neutral or self-contained interaction cues.	Σ CU_augmentation
PL_TBC (Tunable Boundary Coverage)	Number of distinct boundary-related interaction dimensions that can be actively adjusted by the user.	Σ CU_tunable_i where CU_i affects relational boundaries

Design Prediction Parameters

Parameter	Source
DP_P_SP	Designer prediction (Likert)
DP_P_EW	Designer prediction (Likert)
DP_P_HL	Designer prediction (Likert)
DP_R_RA	Designer prediction (Likert)
DP_R_BP	Designer prediction (Likert)
DP_R_SB	Designer prediction (Likert)
DP_R_SO	Designer prediction (Likert)

User CPR measures - Perception

Parameter	Source
P_SP	User perception (Likert)
P_EW	User perception (Likert)
P_HL	User perception (Likert)

User CPR measures - Relation

Parameter	Source
R_RA	User perception (Likert)
R_BP	User perception (Likert)
R_SB	User perception (Likert)
R_SO	User perception (Likert)

Consent Form - Study I

You are invited to take part in a user experience (UX) research study conducted as part of a Master's Thesis at Politecnico di Milano. The aim of this study is to evaluate how designers' choices of anthropomorphic design impacts on users perception and interaction with social AI. Your feedback will contribute to academic research on anthropomorphic design and human–AI interaction. The session will last approximately 30 minutes. You will be presented with several scenario descriptions involving the use of a social AI agent. For each scenario, you will be asked to: select and configure interface elements from a predefined UI component library, optionally include or exclude tunable controls and adjust your decision based on a set of provided design principles guideline. After this, you will be asked to fill out a Likert questionnaire predicting how users might perceive and interact with the system. **There are no right or wrong answers.**

Observation and Recording

- Your interaction with the prototype may be observed for research purposes.
- The session may be audio- or screen-recorded to support later analysis.
- Recordings will be used solely by the researcher and will not be shared publicly.

Confidentiality and Data Use

- All data collected will be treated as strictly confidential.
- No personally identifiable information will be included in any report, publication, or presentation.
- Questionnaire responses will be anonymized and analyzed in aggregate form only.
- Data will be used exclusively for academic research purposes.

Consent Statement

Please confirm the following statements before proceeding:
☐ I have read and understood the information provided above.
☐ I voluntarily agree to participate in this user experience research study.
☐ I understand that my participation is anonymous and that I can withdraw at any time.
☐ I consent to the use of anonymized data for academic and research purposes.
☐ I consent to this session being audio and/or screen recorded for research analysis (optional).

Thank you for your time and participation.
Your contribution is greatly appreciated and plays an important role in this research.

Consent Form - Study II

You are invited to take part in a user experience (UX) research study conducted as part of a Master's Thesis at Politecnico di Milano. The aim of this study is to evaluate how designers' choices of anthropomorphic design impacts on users perception and interaction with social AI. Your feedback will contribute to academic research on anthropomorphic design and human–AI interaction. The session will last approximately 20 minutes, you will be asked to interact with several short scenario-based prototypes. After each interaction, you will be asked to fill out a Likert questionnaires about your experience. You may also be invited to briefly share optional comments about clarity, comfort, or appropriateness of the interaction.

Observation and Recording

- Your interaction with the prototype may be observed for research purposes.
- The session may be audio- or screen-recorded to support later analysis.
- Recordings will be used solely by the researcher and will not be shared publicly.

Confidentiality and Data Use

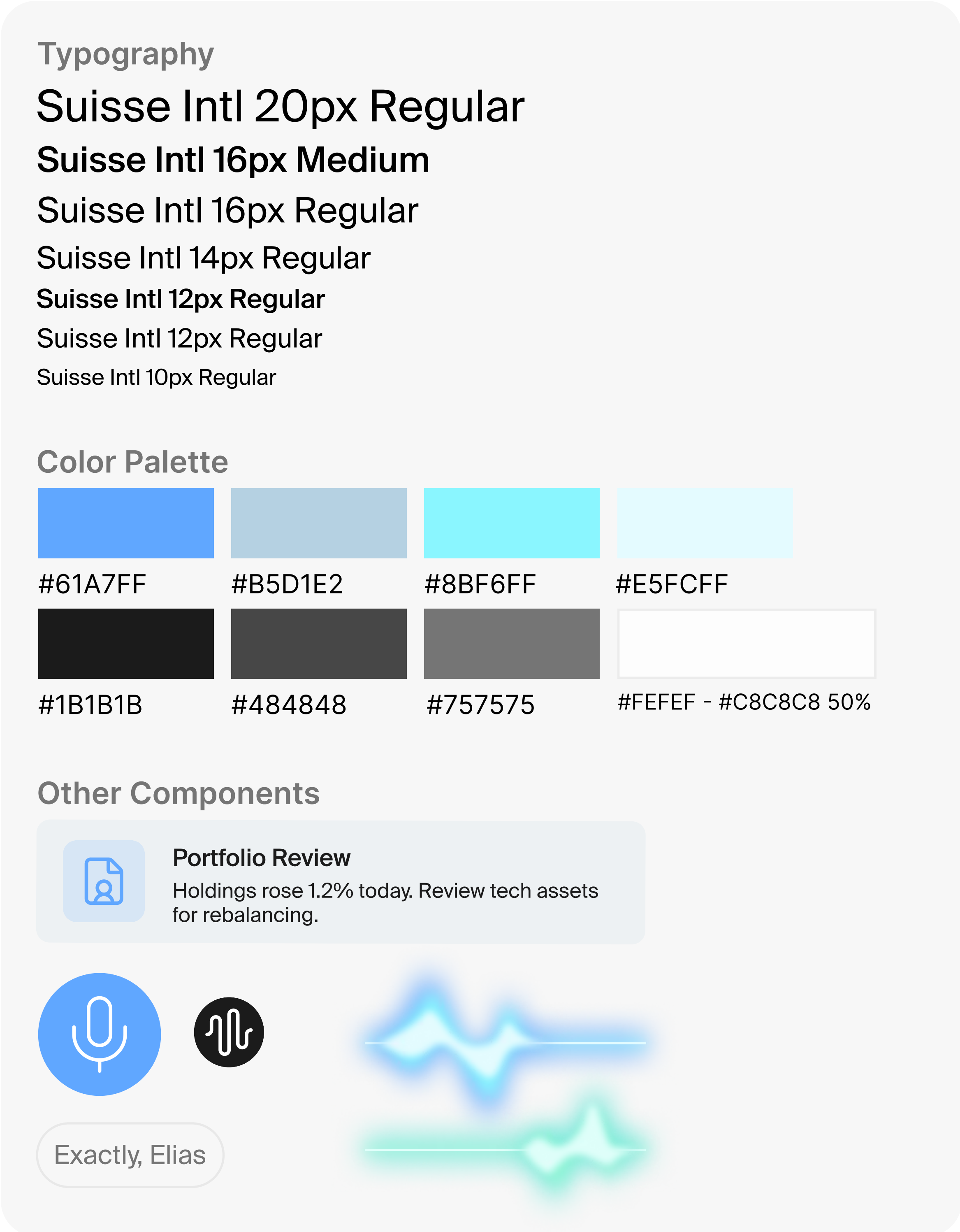
- All data collected will be treated as strictly confidential.
- No personally identifiable information will be included in any report, publication, or presentation.
- Questionnaire responses will be anonymized and analyzed in aggregate form only.
- Data will be used exclusively for academic research purposes.

Consent Statement

Please confirm the following statements before proceeding:
☐ I have read and understood the information provided above.
☐ I voluntarily agree to participate in this user experience research study.
☐ I understand that my participation is anonymous and that I can withdraw at any time.
☐ I consent to the use of anonymized data for academic and research purposes.
☐ I consent to this session being audio and/or screen recorded for research analysis (optional).

Thank you for your time and participation.
Your contribution is greatly appreciated and plays an important role in this research.

Design Systems



Scenario 1_Design System

Typography

Suisse Intl 20px Regular

Suisse Intl 16px Medium

Suisse Intl 16px Regular

Suisse Intl 14px Regular

Suisse Intl 12px Regular

Suisse Intl 12px Regular

Suisse Intl 10px Regular

Color Palette

#FF8D43 #FCEBE0 #B4D8EC #E7F5FD

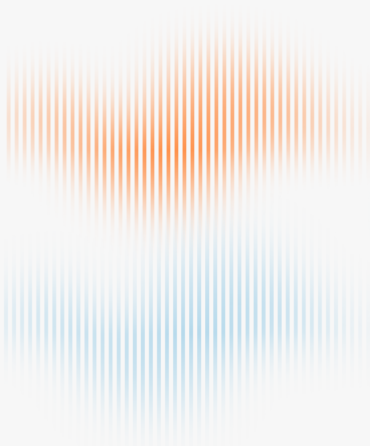
#1B1B1B #484848 #757575 #FEFEF - #C8C8C8 50%

Other Components

Helpful

Appointment on 6th

Thomas, your annual wellness check with Dr. Aris is scheduled for 10:30 AM this Monday at the Central Health Clinic...



Healthy diet

Scenario 2_Design System

Typography

DynaPuff 28px Regular

DynaPuff 20px Regular

Suisse Intl 20px Regular

Suisse Intl 16px Medium

Suisse Intl 16px Regular

Suisse Intl 14px Regular

Suisse Intl 12px Regular

Suisse Intl 12px Regular

Suisse Intl 10px Regular

Color Palette

#FCA6F1	#B5FFFF	#D078E2	#FCA6F1 - #B5FFFF 50%
#1B1B1B	#484848	#757575	#FEFEF - #C8C8C8 50%

Other Components

Roast My Life 🔥

Edit Profile

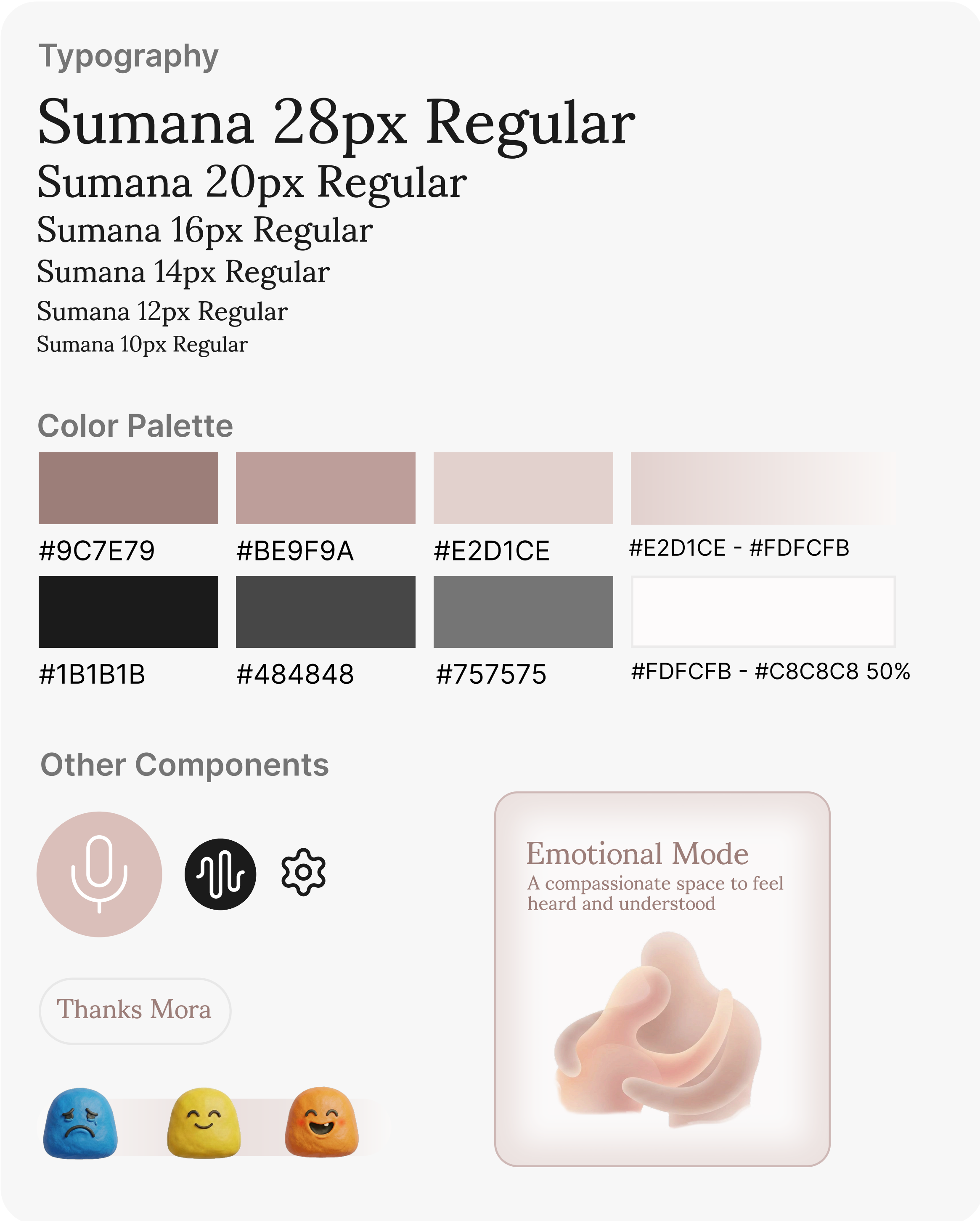
Edit Profile

You get me!

Virtual Idols

Hologram concerts and AI influencers are taking over, get ready to be amazed as you delve into the world...

Scenario 3_Design System



Scenario 4_Design System

Likert - Study I

Based on the anthropomorphic cues you have selected for these scenarios, please predict the typical user's perception and reaction of the agents. Rate each statement on a scale of 1 to 7 to indicate the level of response you expect from the user.

Rating Scale:
• 1 = I predict the user will Strongly Disagree
• 7 = I predict the user will Strongly Agree

For the each agent,

1. Users will feel as if they are interacting with a social entity rather than just a tool.

2. Users will feel the agent warm and considerate.

3. Users will feel the agent's overall presentation and interaction are natural, as if it possessed human-like qualities.

4. Users will perceive the agent as having a distinct identity or character beyond its functional purpose.

5. Rather than just maintaining basic politeness, the users will tend to interact with the agent more as a companion.

6. Users would feel open to sharing their thoughts or individual preferences with the agent.

7. Users would prefer to handle the task through this agent rather than interpersonal contact.

Likert - Study II

Thank you for your participation. In this study, you will evaluate four different AI agents across three interface modes: Greeting, Chatting, and Voice Mode.
Please respond to the following statements based on your subjective experience and perception of each agent.

Rating Scale:
• 1 = Strongly Disagree
• 7 = Strongly Agree

There are no right or wrong answers; we simply value your honest perspective.

For the each agent,

1. During the interaction, I felt as if I were dealing with a living entity rather than just a tool.

2. The agent felt warm and considerate.

3. The overall presentation and interaction of the agent felt natural, as if it possessed human-like qualities.

4. I perceived the agent as having a distinct identity or character beyond its functional purpose.

5. Rather than just maintaining basic politeness, I tend to interact with the agent more as a companion.

6. I felt open to sharing my thoughts or individual preferences with the agent.

7. I would prefer to handle the task in this context through this agent rather than interpersonal contact.

Semi-structured Interview - Study I

1. Can you briefly describe your overall reasoning process when making the design decisions in this task?

2. How did you decide which anthropomorphic cues to include or exclude in your design?
- What factors influenced these decisions?
- Were there any cues you deliberately restrained or toned down?

3. To what extent did you think about the consistency between different cues (such as visual, linguistic, or behavioral elements)?
- Did you imagine the system as having a particular "character" or role?
- If so, how did that influence your design choices?

4. How did you imagine users would relate to the system you designed?
- Did you envision the system as supportive, neutral, or authoritative?
- Were there any relational boundaries you tried to maintain?

5. Beyond the predefined anthropomorphic cues explored in this study, are there other aspects of interaction or system behavior that you think could support anthropomorphic expression and enhance user experience?

Semi-structured Interview - Study II

1. What's your overall feeling and impression of this agent?
- Can you briefly describe your overall experience interacting with this AI agent? At what moments, if any, did the system feel most 'human' to you? Conversely, were there moments when it felt strikingly mechanical or artificial?
- What was the first thing you noticed about the agent's personality?

2. Were there any features that felt "overdone" or made you feel uncomfortable?
- How did the specific design affect your trust and attitude (eg. less open) towards the agent?
- Were there any points where the agent felt too personal or overstepped a boundary in how it interacted with you?

3. If you had to describe this AI agent as a person, what kind of character or personality traits would you assign to it?
- Did you find yourself likely to use social conventions (like saying 'please,' 'thank you,' or using humor) naturally with the agent? Why or why not?

4. Beyond the features you saw today, what other human-like qualities would make this AI agent feel more helpful or engaging for you?

5. If you could change one thing about the agent to improve your experience, what would it be?